Spring 5-15-2021

Essays in Corporate Finance and Machine Learning

Manish Jha
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Finance, Olin Business School

Dissertation Examination Committee:
Todd Gormley, Chair
Alon Brav
Radha Gopalan
Mark Leary
Asaf Manela

Essays in Corporate Finance and Machine Learning

by
Manish Jha

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2021
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am grateful to Radha Gopalan, Todd Gormley, and Asaf Manela for their continuous guidance and support. I also thank Anat Admati, Vikas Agarwal, Taylor Begley, Alon Brav, Mark Chen, Stuart Gillan, Jonathan Kalodimos, Mark Leary, Renping Li, Tao Li, James Pinnington, Paul Pfleiderer, John Puthenpurackal, Janis Skrastins, Pietro Veronesi, and seminar participants at Georgia State University, Midwest Finance Association, Stanford Rising Scholars Conference, Stevens Institute of Technology, University of North Texas, Virginia Tech, Washington University. Computations were performed using the facilities of the Washington University Center for High Performance Computing, which was partially provided through NIH grant S10 OD018091.

Manish Jha

Washington University in St. Louis

May 2021

ABSTRACT OF THE DISSERTATION

Essays in Corporate Finance and Machine Learning

for Arts & Sciences Graduate Students

by

Manish Jha

Doctor of Philosophy in Finance

Olin Business School

Washington University in St. Louis, 2021

Professor Todd Gormley, Chair

My dissertation focuses on two broad questions. First, why do shareholder's preferences vary, and how various agents persuade them? And second, how public perceptions about the financial sector and regulations affect economic outcomes? While my research plan contributes to the two distinct fields of literature, a unifying theme of my research is the use of innovative machine learning techniques to overcome the empirical challenges that would typically prevent measuring these sentiments objectively.

In my Chapter 1, I use a supervised machine learning model on mutual fund family's proxy voting choices to estimate their preferences. I find that hedge fund activists tailor their

campaigns to appeal to the fund families that own a larger share in the targeted firm. Well-tailored campaigns solicit higher engagement and support from the fund families and are more likely to succeed. My findings suggest that activism helps push shareholders' implicit agendas. As per my knowledge, the paper is the first to employ a machine learning model to extract mutual fund preferences, opening up possibilities to analyze other issues where decisions of these institutes matter.

In Chapter 2 with Professor Gormley, we find that mutual fund families conduct more governance research and are less likely to follow proxy advisor recommendations when a firm's bonds represent a larger proportion of their overall portfolio. Our findings suggest that the bond holdings contribute to institutions' incentive to be engaged monitors. In Chapter 3 with Professor Manela and Hongyi Liu, we measure popular sentiment towards finance using a computational linguistics approach applied to millions of books published in eight countries over hundreds of years. We find that the finance sentiment declines after epidemics and earthquakes, but rises following droughts, floods, and landslides. These heterogeneous effects of natural disasters suggest finance sentiment responds differently to the realization of insured versus uninsured risks

Catching the Conscience of Kings: How Activists Pander Mutual Funds

Manish Jha

April 2021

Abstract

Do hedge fund activists tailor their campaigns to pander to mutual fund families? And if so, does the strategy work? Using supervised machine learning on the fund family's proxy voting choices, I estimate their preferences. I find that activists align proxy communications with the preferences of fund families that own a larger share in targeted firms. Activists learn from interactions with fund families and align their campaigns better in subsequent attacks. Tailored campaigns enjoy improved shareholder attention, more votes, and a greater likelihood of success. My findings suggest that activism helps push shareholders' implicit agendas.

JEL Classification: G23, G32, G34

Keywords: Hedge fund activism, Shareholder preferences, Text analysis, Machine learning

Washington University in St. Louis; Campus Box 1133 Simon Hall, St. Louis MO 63130; +1(314) 250 8129; mjha@wustl.edu

> "So the vast majority of companies in the US today are controlled by what I would describe as permanent owners of stock. Think index funds like BlackRock. So the only kind of changes in campaigns we're going to run are ones that benefit the business over decades, and those are the only kind of campaigns you can win. If you have some short-term strategy to make money that's harmful to the company long-term, you're not going to get the support of the BlackRocks, the Vanguards, and the others."
>
> William Ackman, CEO of Pershing Square (NPR 2017)

1 Introduction

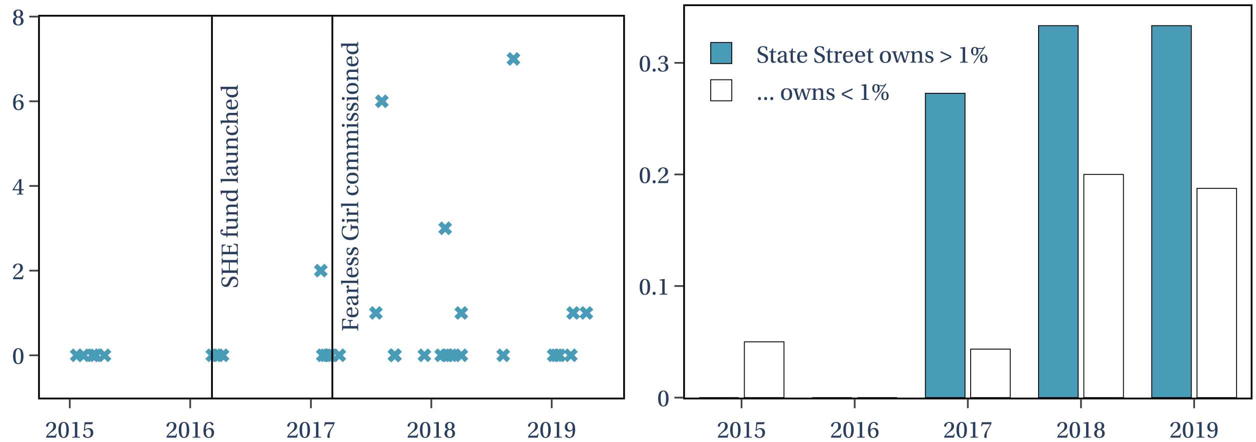
A recent push for social justice and climate change issues, from mutual fund institutions (fund families), has coincided with a shift in hedge fund activists' communications. For example, State Street recently pledged to "vote against the entire slate of incumbent board members if a company does not have at least one woman on its board" (StateStreet 2020). Earlier, the fund family started a gender diversity-focused fund, aptly named SHE, and installed the famous Fearless Girl statue on Wall Street. Simultaneously, activists began using phrases such as "female," "women," etc., in their communications when State Street held more shares in a firm that activists were actively targeting as part of a campaign, illustrated in Figure 1a. In general, investors have become more sensitive to social issues in recent years; and one could argue that phrases discussing gender diversity are bound to show up, irrespective of ownership structure. However, Figure 1b suggests otherwise. Activists highlight gender issues, especially when State Street is a major shareholder. In this paper, I examine whether activists tailor their campaigns to pander to shareholders, and if so, the impact this has on campaigns, tactics, and successes.

Although confrontational proxy fights (attacks) focused on shareholders' preferences occur with significant frequency and attract media attention, research examining interactions

Figure 1:
Activists use gender diversity phrases when State Street is a major stakeholder.
Figure (a) plots the frequency of gender diversity phrases, which include "female," "gender," "woman," and "women" in activists' communications. The procedure used to gather communications is described in Section 2.3. The sample is restricted to attacks where State Street owns more than a percent of the targeted firm's shares. Figure (b) plots the number of attacks that used a gender diversity phrase divided by the total number of attacks in the year. For both the figures, I exclude the 2016 attack on Chico's Clothing, a women's clothing retailer, as the attack text contained references to their business operations, instead of gender diversity.

(a) No. of gd phrases in text, when SS owns > 1%

(b) Fraction of attacks with gender div. phrases



between the parties has been limited. This study fills the gap. Using a comprehensive, hand-collected text dataset of shareholder proposals, I provide the first machine learning-based measure of fund families' preferences. I study whether activists align proxy communications (attack texts) with these preferences and the impact such alignment has on fund families' attention and voting.

Identifying how well an attack is aligned with fund family preferences poses an empirical challenge. Fund families reveal their preferences in many ways, including direct conversations with the targeted firm's (target's) management and board, exiting positions, and proxy voting. The conversations behind the scene and the factors that lead funds to exit positions are not publicly available. As such, researchers have mostly focused on voting record databases to analyze preferences of institutional investors (Bodnaruk and Rossi 2016; Cai, Garner, and Walkling 2009). These voting records, however, do not have the text content of proposals, which restricts researchers from diving into specifics of the issues on which the voting deci-

sions were made. To overcome this challenge, I extend the voting record database by hand-collecting shareholder proposal texts from the Securities and Exchange Commission (SEC) filings over the 2003–2018 period. The extended voting database allows me to analyze voting patterns with respect to the content of the proposals.

To determine an attack's alignment with a fund family's preferences, I define a measure named *Align*, which estimates the family's likelihood of supporting the activist based solely on the activist's communications. First, I relate fund families' voting decisions with shareholder proposals' textual features, using a supervised machine learning model named Support Vector Regression (SVR). The SVR assigns a coefficient to each phrase based on the fund family's voting history. The coefficient of a phrase indicates the marginal increase in the alignment if the text contains one more instance of the phrase. Using frequency and coefficient for each phrase in the attack text, I measure the text's alignment with the fund family.[1]

I find that activists, when designing their attack, focus on the preferences of fund families that own more shares in the target. Activists accomplish this strategy by selectively using phrases that appeal to these institutions. Specifically, a one-standard-deviation increase in target ownership by a fund family, which is approximately 0.63 percentage points relative to an average of 0.09%, is associated with about a 0.4- to 0.7-percentage-point increase in the attack text's alignment with the fund family. The average alignment for a fund family owning ten percent of target shares is 65%, compared to 46% for a fund family with less than 0.01 percent target shares. The increase in support is economically substantial and suggests that activists write their communications to solicit support from larger shareholders.

Using the measure of the attack's alignment, I find that fund families pay more attention to attacks that speak to their preferences. A fund family's attention to an attack, defined as

[1] For instance, during 2016–2017, DWS voted against management in 97% of climate-related shareholder proposals; in contrast, BlackRock, Fidelity, and Vanguard voted against management in 6% of the proposals. I classify proposals as climate-related if the proposal contains any of the following phrases: "climate change," "environmental concerns," "global warming," "renewable energy," "environmental risks." Thus, a SVR model trained on these proposals will figure out that phrases such as "climate change," "environmental concerns," etc., are important for DWS and assign a higher coefficient to these phrases. Subsequently, an attack that focuses on climate-related issues, and thereby uses these phrases, will be considered better aligned with DWS preferences with an *Align* closer to one.

the number of times Internet Protocol (IP) addresses registered under the fund family's name access attack filings on the SEC.gov website, increases with the attack's alignment with the fund family preferences. Specifically, a one-standard-deviation increase in the attack text's alignment, which is approximately 40 percentage points relative to an average of 48%, is associated with a 23% increase in the number of times the fund family access the attack text. The results hold even after controlling for fund family's holdings. For similar holdings in targets, the fund family is more likely to pay attention to the attack that is better aligned with the family's preferences.

I find that activists learn from significant interactions with fund families. I define an attack as a significant interaction between an activist and a fund family if the fund family owns more than one percent of shares in the target. The activists learn from significant interactions and tailor the communications better in subsequent attacks. Specifically, the attack text is 0.9 percentage point more aligned with the fund family preferences, compared to the average 48%, for each significant interaction between an activist and a fund family. The results suggest that activists learn through repeated interactions, which could, in part, explain the increased successes of activists in recent years.

I also find that the activists sway votes when they tailor their text to fund preferences. Fund families often distance themselves from activism in their internal documents, proxy guidelines, and media interviews. For example, internal documents outlining guidelines of Wellington Management, one of Vanguard's outside fund supervisors, call to avoid a public profile unless it benefits clients (WSJ 2019). Nonetheless, a well-aligned text garners higher actual activist support from the fund family. Specifically, among fund families who voted in attack proposals, a one-standard-deviation increase in the attack text's alignment, which is approximately 44 percentage points relative to an average of 56%, is associated with a 0.1-standard deviation increase in the actual activist support, which is three percentage points relative to an average of 53%. Thus, activists gain fund families' votes when they include fund specific issues in their communications.

Along with increased engagement and favorable voting, my results show that attacks that are well-tailored to larger shareholders succeed more often. An attack is deemed successful if the activist's proposals win the shareholder election or the target settles with the activist outside of proxy contests. At the attack level, I measure the use of tailored campaign strategy as the attack text's alignment with fund families aggregated by their ownership coefficients. I find that attacks that employ the pandering strategy are more likely to succeed. Specifically, among attacks that have above-average mutual fund ownership, a one-standard-deviation increase in aggregate alignment, which is approximately 28 percentage points relative to an average of 53%, is associated with an 9.4 percentage points increase in the probability of success for the attack. For reference, activists succeeded in 63% of campaigns over the 2004–2019 period.

To validate the fund family preferences generated from SVR, I check whether the SVR coefficients are in line with their importance in the family's proxy voting guidelines. Mutual funds share proxy voting guidelines with their investors, highlighting factors that they consider while voting proxies. The phrases, which appear more frequently in the proxy guidelines, also have a higher absolute coefficient in the SVR model. The result follows from funds voting consistently on topics they specify in proxy guidelines, pushing coefficients of related phrases in SVR away from zero.

My findings hold subject to various robustness checks. Compared to a dummy text, stitched sequentially from parts of all the attack texts, the actual attack text garners higher aggregate alignment. The difference is significant at 1%. The attacks have become more aligned with the shareholders over the sample period. On average, the attack text garners 53% aggregate alignment, compared to 28% for the dummy text. The 25 percentage points increase in aggregate alignment is associated with 8.4 percentage points higher probability of an activist succeeding or settling the attack. The results also hold for using a straightforward non-machine learning approach. I manually classify shareholder proposals into 25 types, based on their headings, and analyze fund families voting patterns in each type for two years prior to the attack. In their campaigns, the activists include types of proposals that are voted on favorably by the larger

shareholders. Lastly, the results are robust to various specification choices.

Overall, this paper contributes to the literature that studies investor activism. A rich literature studies the characteristics of activists, shareholders, and targets, and the implications of activism for shareholder value and other corporate outcomes.[2] However, relatively little is understood about the interactions between the parties involved, how investors choose their tactics, and what factors contribute to their success. I show that campaign tailoring is an effective way for activists to collectively engage, enabling the small blockholders to govern via voice (Brav, Dasgupta, and Mathews 2019). This paper adds to the literature by showing that shareholders' preferences dictate tactics employed by activists, the issues they fight on, the engagement and support they get, and, ultimately, the outcomes of their campaigns.

My findings supplement Appel, Gormley, and Keim (2019), who also show that firms' ownership structures play a role in determining the choice of tactics by activists. The granularity of the ownership structure in their paper is restricted to the type of institution, such as active or passive institution. In contrast, I identify preferences at a more granular ownership structure based on individual fund families. Institutional investors, even those that belong to the same institution type, have differing ideas of what constitutes the correct course of action.[3] Thus, the activist's strategy depends not only on the types of institutions, which make up the ownership structure, but also to the specific fund families that make up institution types. Furthermore, this paper extends their work by exploring the effects of campaign tailoring on fund families' attention and voting decisions.

This paper also contributes to the strand of literature that explores voting decisions in confrontational proxy contests. Related works show that investors, who are connected with activists, vote against targets more often (He and Li 2018), and activists regularly interact with

[2] For characteristics of parties involved, see Bebchuk, Cohen, and Hirst (2017); Bradley, Brav, Goldstein, and Jiang (2010); Brav, Jiang, Partnoy, and Thomas (2008); Clifford (2008); Greenwood and Schor (2009); Gu and Zhang (2020); Mietzner and Schweizer (2014). For the impact of activism, see Boyson, Gantchev, and Shivdasani (2017); Brav, Jiang, and Kim (2015a); Brav, Jiang, Ma, and Tian (2018).

[3] Matvos and Ostrovsky (2010); Morgan, Poulsen, Wolf, and Yang (2011) observe systematic differences in mutual fund voting, indicating divergent preferences. For example, in Trian's proxy fight with Proctor and Gamble, Vanguard sided with the target while BlackRock and State Street voted with the activist, even though they fall in the same type - passive institutions.

asset managers about their plans for a target firm (Edmans and Holderness 2017). Brav, Jiang, Li, and Pinnington (2018) show that the fund families with high variation in the votes cast over time could be persuaded to vote for the activist. Other research finds that mutual funds support target's management when they have business ties (Ashraf, Jayaraman, and Ryan 2012; Cvijanović, Dasgupta, and Zachariadis 2016; Davis and Kim 2007), or cross holdings (Harford, Jenter, and Li 2011; Matvos and Ostrovsky 2008). Factors, such as governance failures at mutual funds (Chou, Ng, and Wang 2011) or a common educational background between fund managers and the company's CEO (Butler and Gurun 2012), also add to target friendly voting. This paper supplements the existing literature by showing that what activists say and the issues they raise also affect shareholders' voting decisions.

Finally, my paper contributes to a growing body of work that applies text-based analysis to fundamental economic questions, in this case quantifying shareholder preference. Prior works employ a more manual approach, classifying proposals into different classes based on issues raised, sponsoring institutions, etc., and subsequently assessing fund family voting.[4] Two recent papers that employ statistical techniques to quantify shareholder preference include (i) Bubb and Catan (2018), who undertake a principal components analysis to classify mutual funds in terms of how they follow distinctive philosophies of corporate governance, and (ii) Bolton, Li, Ravina, and Rosenthal (2020), who employ a scaling application to place fund families into social-orientation and profit-orientation dimensions. However, the preferences in these papers are based on mutual fund voting patterns with respect to each other and do not take into account the underlying content of the proposals. This paper is the first to extract proposal's contents and employ a supervised machine learning model to extract mutual fund preferences. The model is interpretable, in the sense that coefficients of phrases reflect their relative importance, and can focus on narrow variations in preferences at fund family

[4]Related literature shows that proposals sponsored by institutions get substantially more support, compared to proposals sponsored by individuals (Gillan and Starks 2000), that less myopic funds are more likely to vote for environmental and social issues (He, Kahraman, and Lowry 2018), and that holdings-based corporate social responsibility score for funds is positively associated with voting favorably on social responsibility proposals (Li, Patel, and Ramani 2019).

level across time.

2 Ownership, shareholder proposal, and activism data

2.1 Fund family holdings in target

I use the Center for Research in Security Prices (CRSP) mutual fund holdings data, available via Wharton Research Data Services (WRDS), to compute mutual fund holdings in target stock as a percentage of its market cap. Since 2003, all open-ended mutual fund and ETF portfolios are required by the SEC to file their holdings quarterly. I calculate each stock's total market cap using the CRSP monthly file as the sum of shares outstanding multiplied by price for each class of common stock associated with a firm. For private firms, such as Dell Technologies, for which share price is not available, I use the book value of common equity from S&P Capital IQ as the market cap. In total, I gather holdings information for 438 attacks over the 2004–2019 period. Section 2.3 explains the method to identify the attacks. On average, mutual funds hold 18.5% of target stocks, with the median being 17.6%. In contrast, the median activist stake in the target before an attack ranges from 6.3% to 8.8% (Boyson and Mooradian 2011; Brav, Jiang, and Kim 2010, 2015b).

To aggregate individual funds at the institution level, I manually match funds to the larger fund family using their name, while also accounting for subsidiaries within each institution. For example, Allianz Global Investors purchased both Nicholas-Applegate Capital Management and Pacific Investment Management Company in 2000, and in 2008, it invested $2.5 billion in Hartford Financial Services Group. So, I assign all funds with names containing "Allianz," "Nicholas-Applegate," "PIMCO," and "Hartford" to the Allianz fund family. When aggregating positions to the fund family level, I exclude fund-level positions with a negative value, reflecting short positions. Subsequent findings are similar if I instead keep these negative positions or use their absolute value when aggregating. Out of the 438 attacks, 46 have a fund family with more than ten percent target shares. A further 142 attacks have at least one

fund family that owns more than five percent of target stocks. These large chunks of voting blocks concentrate the diffused share votes and are often the precursor to facilitating change (Alchian and Demsetz 1972; Bainbridge 2005; Shleifer and Vishny 1986).

Figure 2a illustrates the list of largest fund families invested in targets at the initiation of an attack. Vanguard is the largest shareholder in 155 attacks, followed by Fidelity at 58, and BlackRock at 52. The list of largest shareholders contains fifty unique mutual fund families. It seems like the usual suspects such as BlackRock, Fidelity, and Vanguard are the major shareholders in all the attacks; and activists have to simply pander to them irrespective of the attack. However, Figure 2b shows otherwise. The top three fund families are the largest shareholders in 61% of the attacks. The distribution of holdings for the top three fund families demonstrates that these fund families do not play a significant part in many of the attacks. BlackRock, Fidelity, and Vanguard own less than 1% target share in 47%, 65%, and 36% attacks, respectively. The distribution underlines that the institutes holding voting power vary across targets. As such, the activists have to tailor their approach for each attack, instead of catering to the same few fund families across attacks.

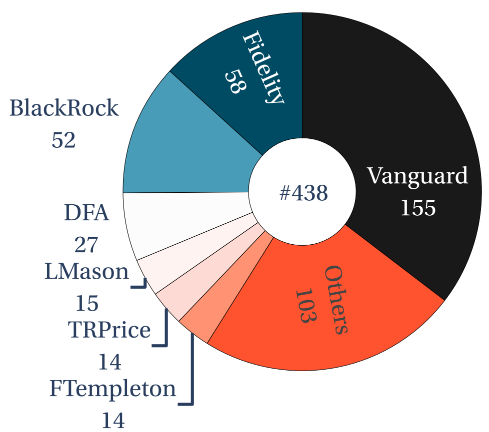
2.2 Shareholder proposal text

I obtain data on fund families' voting from Voting Analytics, which is compiled by ISS. The database includes mutual fund proxy voting records (N-PX filings with SEC). Since July 2003, the SEC has required mutual funds to disclose how they vote proxies. As such, proxy voting data in the paper covers the period from July 2003 to October 2018. The dataset contains proposals sponsored by the firm's management as well as shareholders. I exclude management-sponsored proposals, as voting on these proposals is largely perfunctory and less revealing of fund family preference. Focusing on the shareholder proposals also helps with the imbalanced dataset problem, which occurs if the training data contains many more samples from one class than from the rest of the classes.[5]

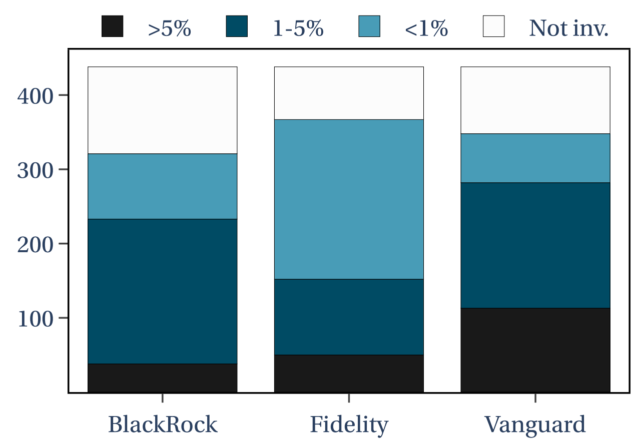
[5]Kubat, Matwin, et al. (1997) show that adding examples of majority class could have a detrimental effect on the learner's behavior. For an average proposal (including management sponsored proposals), mutual fund

Figure 2:
Fund families that own significant voting power vary across attacks.
Figure (a) plots fund families with the largest stock ownership in the targets at the initiation of attacks over the 2004–2019 period. The holdings data is gathered from CRSP, and aggregated to parent institutions that manage these funds. The total number of attacks is annotated at the center. Figure (b) plots the distribution of investment in stocks as a percent of market cap across attacks for BlackRock, Fidelity, and Vanguard.

(a) Largest fund families in the target



(b) Number of attacks, at different holdings



Next, I add text to the proposals in the voting database. Firms are required to file the Definitive Proxy Statement Form (DEF 14A) with the SEC when a shareholder vote is needed. I match shareholder proposal voting information with DEF14A available via the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. I use the Central Index Key (CIK) to the Committee on Uniform Securities Identification Procedures (CUSIP) link table, provided by WRDS, to match the two datasets. I supplement this link table with the CIK-CUSIP database, made from parsing 13D/13G filings (Schwartz-Ziv and Volkova 2020). To match the DEF14A proposal text with the voting database, I use: (i) text-similarity of proposal's heading, (ii) proposal sequencing number, and (iii) filing date. In total, I assign 6,176 proposal texts to the shareholder proposals in the voting database. Appendix A.1 explains the process of extracting shareholder proposals and matching them with voting records in detail.

Table 1 reports the number of shareholder proposals with matched text for each year from

families are unlikely to vote against the management. Over the 2003–2018 period, mutual funds voted against management recommendations for only 9% of management sponsored proposals, as opposed to 42% for shareholder sponsored proposals.

Table 1:
The largest fund families tend to follow management recommendations
This table reports fund family voting on shareholder proposals by year. The sample contains all the shareholder proposals for which the text was available from the SEC. In Column (1), the number in parentheses indicates the percent of proposals for which ISS recommended against the management. Columns (2), (3), (4), (5), and (6) show the voting history of the five largest US mutual fund families by asset under management (RelBanks 2017). The number inside parentheses indicates the percent of proposals with an against management vote during the year.

Year	Shareholder Proposals (1)	BlackRock (2)	Charles Schwab (3)	Fidelity (4)	State Street (5)	Vanguard (6)	Full Sample (7)
2003	42 (40)	24 (27)	15 (20)	-	-	-	841 (29)
2004	499 (45)	439 (27)	343 (16)	15 (42)	7 (14)	17 (34)	15,692 (31)
2005	480 (50)	452 (35)	380 (16)	400 (26)	337 (10)	425 (40)	24,777 (31)
2006	488 (62)	440 (39)	483 (57)	479 (24)	83 (16)	483 (19)	37,437 (42)
2007	398 (60)	364 (47)	384 (62)	344 (27)	46 (11)	380 (18)	27,674 (41)
2008	374 (62)	261 (52)	330 (65)	334 (37)	324 (21)	313 (26)	21,228 (45)
2009	529 (72)	406 (56)	510 (68)	523 (40)	494 (30)	459 (16)	39,973 (52)
2010	360 (77)	344 (33)	301 (49)	327 (36)	317 (22)	327 (20)	24,980 (51)
2011	257 (78)	249 (45)	193 (63)	226 (46)	221 (36)	226 (34)	14,598 (59)
2012	358 (65)	328 (41)	242 (49)	329 (39)	304 (42)	292 (31)	21,577 (51)
2013	497 (65)	492 (27)	412 (37)	494 (23)	464 (35)	478 (18)	34,142 (44)
2014	534 (62)	530 (20)	418 (30)	521 (22)	516 (39)	521 (17)	38,383 (41)
2015	573 (71)	538 (31)	476 (17)	540 (22)	533 (38)	533 (15)	51,438 (44)
2016	413 (66)	391 (24)	360 (19)	389 (19)	380 (37)	393 (17)	34,507 (41)
2017	371 (60)	353 (27)	296 (25)	346 (28)	331 (30)	355 (20)	30,497 (40)
2018	3 (100)	2 (0)	-	2 (0)	2 (0)	3 (0)	104 (41)
Total	6,176 (63)	5,613 (35)	5,143 (40)	5,269 (29)	4,359 (31)	5,205 (21)	417,848 (44)

2003 to 2018. Overall, my sample contains 359 unique mutual fund families, with an average (median) of 1,163 (479) voting observations over the 2003–2018 period. The ISS recommended against the management for 63% of shareholder proposals. As expected, prominent fund families are well diversified and have voted in almost all the shareholder proposals in any particular year. BlackRock voted on 91% of shareholder proposals between 2003 and 2018. Among the top five largest US mutual fund families by assets under management, Vanguard follows management recommendations the most, followed by Fidelity and State Street. The larger fund families, compared to ISS or the overall sample, are less likely to vote against the management.[6] In contrast, smaller fund families, which are often active and follow proxy advisor

[6]Related literature shows that fund families support management when they have business ties (Cvijanović, Dasgupta, and Zachariadis 2016; Davis and Kim 2007), cross-ownership (Hemphill and Kahan 2019; Matvos and

recommendations (Ertimur, Ferri, and Oesch 2013; Iliev and Lowry 2015), are more willing to show their dissent.

2.3 Attack text

During an attack, activists and targets put forth their viewpoints to shareholders and send proxy cards to solicit votes. The shareholders sign and return proxy cards to the party they support. Both parties accumulate votes via the returned proxy card and use them at the shareholders meeting. The communications often include a letter to shareholders, which discusses activist's rationale for the attack. I combine these letters to create the attack text. Activists file a Preliminary Proxy Statement in Connection with Contested Solicitations (PREC14A) and Definitive Proxy Statement in Connection with Contested Solicitations (DEFC14A). Activists also file Additional Definitive Proxy Solicitation Materials Filed by Non-Management (DFAN14A) if the registrant does not support the proxy solicitations. The forms are available to the public via the SEC's EDGAR system. I parse each DFAN, DEFC, and PREC filing (attack filing) to extract the filer and subject company. I restrict the sample by cross-referencing the filer with a list of investment managers that have filed a Schedule 13F holdings report, a requirement for institutions holding more than $100 million in US stocks at some point in their history.[7]

I start my attack text dataset from 2004, six months after the mutual fund voting records are available, to have at least a hundred mutual fund voting records for constructing preferences. I bunch together all the filings for a filer-subject pair if the difference between consecutive filing dates for these documents is less than 180 days. I get a total of 533 confrontational proxy

Ostrovsky 2008; Xie and Gerakos 2020), other peers supporting management (Matvos and Ostrovsky 2010), or pension ties (Ashraf, Jayaraman, and Ryan 2012). On the other hand, proxy advisors, such as ISS and Glass Lewis, often recommend against management to justify their employment (Allaire 2013). Fund families support management, even when it seems to go against shareholders' interests. In 2015, BlackRock, State Street, and Vanguard voted against Trian's candidates for DuPont's board. After the failed activism event, Dupont's stock price declined abruptly, the company missed earnings, and the CEO "retired voluntarily." Shortly after that, DuPont announced a merger with Dow Chemical to handle problems that Trian had identified.

[7] Alexander, Chen, Seppi, and Spatt (2010); Brav, Jiang, Li, and Pinnington (2018); Fos and Tsoutsoura (2014); Norli, Ostergaard, and Schindele (2014) employ a similar approach to identify attacks. Greenwood and Schor (2009) also use 13F to exclude corporate cross-holdings with activism from portfolio investors.

Table 2:
Only a few activists have led more than ten attacks
This table reports prominent lead activists and firms they targeted over the 2004–2019 period. Some of the larger target firms, in terms of their market cap, are also listed.

Activists (# attack)	Major Targets
Breeden Cap Mgmt (2)	Applebees, PIMCO
Icahn Enterprises (39)	Time Warner, Yahoo, Dell, eBay, AIG, Clorox, Family Dollar, Motorola, Tyson, Xerox, Cigna, Biogen
Land & Buildings Inv Mgmt (8)	Macerich, MGM Resorts, Taubman Centers
Marcato Cap Mgmt (3)	Lear Corp, Deckers Outdoor Corp, Buffalo Wild Wings
P Schoenfeld Asset Mgmt. (3)	T-Mobile
Pershing Square Cap Mgmt (2)	Allergan Inc, Automatic Data Processing
Starboard Value (22)	Bristol Myers, Office Depot, Dollar Tree, Yahoo, AOL
Steel Partners (11)	Rowan Companies, GenCorp
Third Point (9)	Dow Chemicals, Campbell, Yahoo
Trian Funds (3)	P&G, DuPont, Heinz

attacks over the 2004–2019 period, with an average of 7.9 (median 5) filings per attack. The number of attacks involving proxy contests is significantly lower than the general hedge fund campaigns.[8] In Appendix A.2, I explain the method and textual cues that I employ to extract activists' communication with shareholders.

Table 2 shows notable activists, along with their attacks. Some of the well-known activists such as Icahn Capitals with 39 attacks, and Starboard Value with 22 attacks, lead the list. Nonetheless, the activism share is fragmented, and only nine activists have double-digit attacks over the 2004–2019 period. The 533 attacks are shared among 177 unique activists. Often a group of attackers together target a firm as a wolfpack (Coffee Jr and Palia 2016; Wong 2019) or co-ordinate by co-filing Schedule 13Ds (about 22% of Brav, Jiang, Partnoy, and Thomas (2008) sample). The lead attacker files attack documents with the SEC, and only its name appears in the "filed by" section of the document. Thus, the share of each attacker is higher than shown in Table 2.

[8] Activists only use proxy contests as a threat since it is costly for both parties, and only around 10-12% of hedge fund campaigns threaten a proxy contest (Gantchev 2012). In activism literature, 13D filings are often used to identify hedge fund campaigns. These are beneficial ownership filings, required for investors when they own more than 5% in any class of a firm's securities and intend to influence the firm. I do not include 13D filings in attack text, as they often do not contain information related to activists' contentions with the manager. I also do not use media reports to determine the contention, as the sources and linguistic differences add more noise than information.

Activists discuss various issues in these communications, including reasons for management's failure, the areas to focus for the firm, and plans to improve shareholder value. The 2017 Trian Fund Management attack on Proctor & Gamble (P&G) illustrates these discussions. Trian started its activist campaign against P&G for its nominee, Nelson Peltz, to the board of directors at the 2017 annual shareholders meeting. P&G responded that its board and management team is actively executing its strategy to achieve balanced, sustainable long-term growth and value creation. Trian Fund mailed a letter to shareholders detailing why it views that adding an independent director can lead to the breakthrough ideas P&G needs, and why it is necessary to cut through P&G's rhetoric so that shareholders can make an informed decision. In total, Trian filed a total of 60 DFAN14A, which I combine to create the attack text for this attack. While Nelson Peltz was not elected, the board later decided to expand and accommodate him.

3 Methodology and descriptive statistics

I exploit fund family voting on shareholder proposals to measure investor preference. McCahery, Sautner, and Starks (2016) report that institutional investors frequently employ voice in their engagements. 53% of the institutional investors report voting against management as a shareholder engagement channel, second only to discussions with top management (63%). I prefer the voting channel over engagement behind the scenes because of the public availability and standard nature of the voting data points. I assume throughout that the voting choice by the fund family provides a good and stable reflection of its concerns.[9]

[9] Aggarwal, Saffi, and Sturgess (2015) note that investors recall loaned shares, thus incurring a financial cost, ahead of the proxy record date to exercise voting rights, and use the proxy process to affect corporate governance. The SEC (2019) directive expects fund managers to implement a reasonably formulated voting policy and cast votes consistent with its voting policies and procedures.

3.1 Training SVR on shareholder proposals

To standardize the different ways fund families vote against the management for a proposal, I define a dummy *Align*, which indicates the level of alignment between fund family preferences and proposal text. For a particular mutual fund portfolio, *Align* is one if the portfolio does not precisely follow management's recommendation for the proposal. For example, *Align* is one if the management recommendation is "for," and the mutual fund portfolio votes "abstain," "do not vote," "withhold," or "against." To get fund family level *Align*, I average portfolio level *Align* across the fund family for the proposal.

I also standardize shareholder proposal texts by removing non-English words, stop words, case, HTML tags, punctuation, digits, inflectional endings, and filler words. I use n-grams of up to five-word phrases to extract features from the text. I omit phrases that appear in less than 1% or more than 70% of the proposals to remove misspelled and frequently used legal terms. I get a total of 9,832 phrases, comprised of 2,465 unigram, 4,991 bigram, 1,593 trigram, 567 4-gram, 216 5-gram. Each shareholder proposal's text is, therefore, represented by $\boldsymbol{x}_s$, a K = 9,832 vector of phrases frequencies, where $x_{p,s}$ = count of phrase p in shareholder proposal text, s.[10] I analyze how the fund family voted on the proposal's text, using a linear regression model:

$$Align_{s,f} = \alpha_f + \boldsymbol{\beta_f} \cdot \boldsymbol{x}_s + \nu_{s,f} \tag{1}$$

where $Align_{s,f}$ is the fraction of funds (managed by fund family, f) that voted against management's recommendation for the shareholder proposal, s. $\boldsymbol{\beta_f}$ is a K vector comprised of a coefficient for each phrase. Predicting the text's alignment with fund family preferences is a regression problem like any other. However, the high dimensionality of the text makes or-

[10]Lemmatization removes inflectional endings and returns the base or dictionary form of a word. So, "stockholders will be asked to approve a proposal" becomes "stockholder will be ask to approve a proposal." I remove stopwords, identified from Python's natural language toolkit English corpora, which shortens the sentence to "stockholder ask approve proposal." Finally, countvectorizer, available via scikit-learn, converts "stockholder ask approve proposal" to four uni-grams – "stockholder," "ask," "approve," "proposal;" three bi-grams, and so on.

dinary least squares and other standard techniques infeasible. To circumvent the problem, I employ a supervised machine learning model - SVR, developed by Drucker, Christopher, Kaufman, Smola, and Vapnik (1997). The method is used in finance literature by Kogan, Levin, Routledge, Sagi, and Smith (2009) to predict risk from financial reports, and Manela and Moreira (2017) to measure news implied volatility. The SVR estimation procedure performs well for short samples with a large feature space K. While the full treatment of SVR is beyond the scope of this paper, I document an intuitive glimpse into this method and the structure that it implicitly imposes on the data. SVR minimizes the following objective:

$$H(\boldsymbol{\beta_f}, \alpha_f) = \sum_{t \in train} g_\epsilon(v_{s,f} - \alpha_f - \boldsymbol{\beta_f} \cdot \mathbf{x}_s) + \frac{\boldsymbol{\beta_f} \cdot \boldsymbol{\beta_f}}{2c} \tag{2}$$

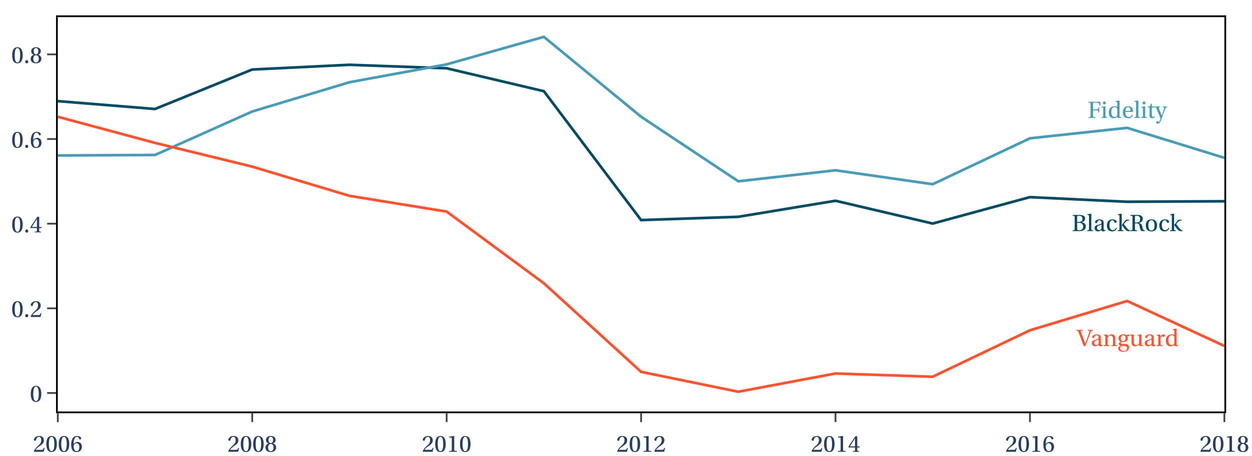
where $g_\epsilon = \max(0, |e| - \epsilon)$ is an ϵ-insensitive error measure, which ignores errors of size less than ϵ, i.e., it only penalizes samples whose prediction is at least ϵ away from their true value. The loss function is similar to Gaussian linear regression, except SVR adds a penalty for each dimension of coefficient, $\boldsymbol{\beta_f}$, that deviates from zero via an inverse regularization parameter c. I use ϵ insensitive zone value of 0.001, and inverse regularization parameter c = 0.0001. Appendix B describes the process in more detail. The minimizing coefficients vector $\boldsymbol{\beta_f}$ can be completely described as a linear combination of the training observations. Only some of the training observations' coefficients are non-zero, and the associated data values are called the support vectors, thus the name support vector regression. At the end of the process, SVR assigns coefficients to each of the phrases used in shareholder proposals, where a positive (negative) coefficient implies that the phrase increases (decreases) the text's alignment with fund family preferences.

Figure 3a plots variations in the coefficient of "simple majority vote" for BlackRock, Fidelity, and Vanguard across time. A simple majority standard represents that half (plus one) of all votes cast should be cast "for" to consider a ballot item passed. Most fund families prefer this standard over the super-majority standard, which specifies a higher threshold than one half. The SVR coefficients are interpretable, meaning one could infer from the graph that Fi-

Figure 3:
SVR coefficients are interpretable and rooted in proxy voting choices.
Figure (a) plots SVR coefficients for "simple majority vote" calculated on December 31st of each year for BlackRock, Fidelity, and Vanguard. The calculations are based on the fund family's proxy voting choices in shareholder proposals during the two years before the calculation date. The coefficients indicate the marginal increase in the text's alignment with fund family preferences if the text contains one more instance of the phrase. For example, a coefficient of 0.004 for Fidelity in December 2010 indicates that Fidelity is 0.4 percentage points more likely to vote against the management for every instance of "simple majority vote" in the proposal text. Figure (b) plots the fraction of shareholder proposals containing "simple majority vote," where the fund family voted against the management recommendation. The proposals are gathered from DEF14A filings, available via EDGAR.

(a) SVR assigned coefficient for "simple majority vote"

(b) The fraction of shareholder proposals containing "simple majority vote," where the fund family voted against the management recommendation



delity is more likely to vote against management if the proposals include measures to install simple majority standards. The coefficients are also rooted in voting patterns of fund families. Figure 3b illustrates that Fidelity has indeed voted against management when the phrase "simple majority vote" is mentioned in the proposals. Thus, the text-based preference has two useful attributes: (i) it allows fund preference to be time-variant, dependent on the way funds voted in the two years before the measurement date. (ii) its variation is interpretable and provides insight into issues that are important to the fund family. The cost of SVR is that the kernel cannot adapt itself to concentrate on sub-spaces of $\mathbf{x}$ (Hastie, Tibshirani, and Friedman 2009). For example, if a fund family is environmentally conscious and votes against management when specific phrases such as "climate change" occurs in proposal text, the SVR will assign a high positive coefficient to "paris." Even though the phrase "paris" is orthogonal to voting decisions in most cases, it gets a positive coefficient from co-occurrence with "climate change" in "Paris Agreement on Climate Change." Ultimately, how well a machine learning method works is dependent on how well it predicts out-of-sample observations. The out-of-sample mean absolute error for *Align* for an average fund family is 0.40. SVR (using coefficients in the proposal text) reduces the mean absolute error to 0.24.

3.2 Measuring the attack text's alignment with fund family preferences

To determine fund family preferences for a particular attack, I analyze its voting choices on the shareholder proposals from the first attack filing date (attack date) to two years before the attack date. To have sufficient training samples, I only consider families that have voting information for at least a hundred proposals during the period. For example, in the P&G 2017 attack by Trian Funds, the first proxy (between DEFC, DFAN, and PREC) filed by Trian was a DFAN14A on July 17th, 2017. So, I consider shareholder proposals during the July 17th, 2015–July 17th, 2017 period. A total of 189 fund families voted in at least a hundred of these shareholder proposals.

For each fund family, I solve SVR regression Equation 3, which assigns a coefficient to each

of the phrases in shareholder proposals. Using frequencies of phrases in the attack text and their associated coefficients, I estimate the attack text's, a, alignment with fund family, f, preferences or $\widehat{Align}_{a,f}$ as:

$$\widehat{Align}_{a,f} = \widehat{\alpha}_f + \widehat{\boldsymbol{\beta}}_{\boldsymbol{f}} \cdot \boldsymbol{x}_a \tag{3}$$

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{f}}$ is the estimated K-vector containing coefficients assigned to phrases derived from training the SVR model on the fund family's voting history. The attack text is represented by **x**, a K-vector of phrases' frequencies in the attack text. The estimated alignment, or $\widehat{Align}$, is the likelihood of fund family supporting the activist based solely on attack text and is bounded by zero and one. Thus, if an attack text uses phrases such as "simple majority vote," that are important to Fidelity, indicated by the phrase's positive coefficient in Fidelity's SVR model, the attack text's alignment with Fidelity preferences will be closer to one.

3.3 The phrases that matter

SVR coefficients vary across fund families, as well as across time. Table 3 lists the top ten phrases associated with increased alignment with BlackRock, Fidelity, and Vanguard on December 31st, 2017. The list is filtered for phrases that contain at least three words to provide context. Classification of common stocks is an important issue to all the three fund families. BlackRock submitted a petition to the New York Stock Exchange and Nasdaq to require companies to eliminate unequal voting rights enshrined in different share classes (WSJ 2018). Likewise, Fidelity and Vanguard mention in their proxy guidelines that they "generally support proposals to recapitalize multi-class share structures," and "are opposed to dual-class capitalization structures that provide disparate voting rights" (Fidelity 2019; Vanguard 2018).

The three fund families also differ in their relative importance to specific issues. For BlackRock, executive compensation and disclosure are important. During 2016–2017, BlackRock voted against the management in 30% of the shareholder proposals that contain the word

Table 3:
The importance of phrases in voting decisions varies across fund families
This table reports phrases associated with an increase in the alignment of attack text with fund family preferences. The list is based on fund families' preference on December 31st, 2017; thus, the SVR considers shareholder proposal voting patterns over the 2016–2017 period. The phrases are stripped of cases, punctuations, stopwords, and noun/verb forms. The list is filtered for phrases that contain at least three words.

BlackRock	Fidelity	Vanguard
class common stock	class common stock	proxy access proposal
vote per share	vote per share	vote per share
incentive stock option	simple majority vote	class common stock
executive compensation program	corporate political contribution	director executive officer
recommend vote proposal	special meet proposal	stock option award
include proxy material	please vote protect shareholder value	new independent director
statement satisfy bylaw	name executive officer	name executive officer
statement satisfy bylaw applicable	please vote protect	stock per share
disclosure statement satisfy	please vote protect shareholder	board director board
disclosure statement satisfy bylaw	vote protect shareholder value	enhance shareholder value

"disclosure." In comparison, the number is 14% for Fidelity and 12% for Vanguard. For Fidelity, "simple majority vote" and "corporate political contribution" are the top phrases. Fidelity's proxy guidelines mention that they will oppose anti-takeover provisions, including super-majority voting (Fidelity 2019). For Vanguard, proxy access proposals, which call for the opportunity to place director nominees on a company's proxy ballot, are important and are also mentioned in its proxy guidelines (Vanguard 2018).

The phrases in Table 3 seem similar to each other. For example, the bottom three phrases for Fidelity are related to protecting shareholder value. When sorted by coefficients, the bunching of phrases occurs because similar phrases are associated with almost identical voting behavior by the fund family. Inline, the SVR assigns close coefficients to these phrases. Thus the K (9,832) phrase vector has phrases referring to the same issues bunched together. Table 3 shows the top part of the larger sample.

Table 4:
Summary statistics
This table reports summary statistics of key variables: (i) Holding, which is the fraction of the target's equity owned by a fund family, and (ii) $\widehat{Align}$, which is the predicted attack text's alignment with fund family preferences. Columns (1) and (2) include results for the fund families that have voted in at least a hundred shareholder proposals in the two years before an attack. Columns (3) and (4) report the numbers for a smaller sample of fund families that hold shares in the target.

	All fund families		…invested in target	
	Holding (in %)	$\widehat{Align}$	Holding (in %)	$\widehat{Align}$
Observations	66,836	66,836	12,582	12,582
Mean	0.09	0.48	0.49	0.48
Std. Deviation	0.63	0.40	1.37	0.40
Minimum	0	0	0	0
25th Percentile	0	0.05	0.01	0.05
Median	0	0.44	0.06	0.43
75th Percentile	0	0.94	0.26	0.97
Maximum	18.58	1	18.58	1

3.4 Sample and descriptive statistics

For my primary analysis, I use the attack text's alignment with each fund family who has voted in at least a hundred shareholder proposals in the two years before the attack date. Some of these fund families are not invested in the target; in which case, I assign an equity share of zero. My sample includes 522 attacks (438 with at least one invested fund family), involving 287 unique fund families, over the 2004–2019 period. In total, my sample contains 66,836 (12,582 with non-zero holdings) observations.

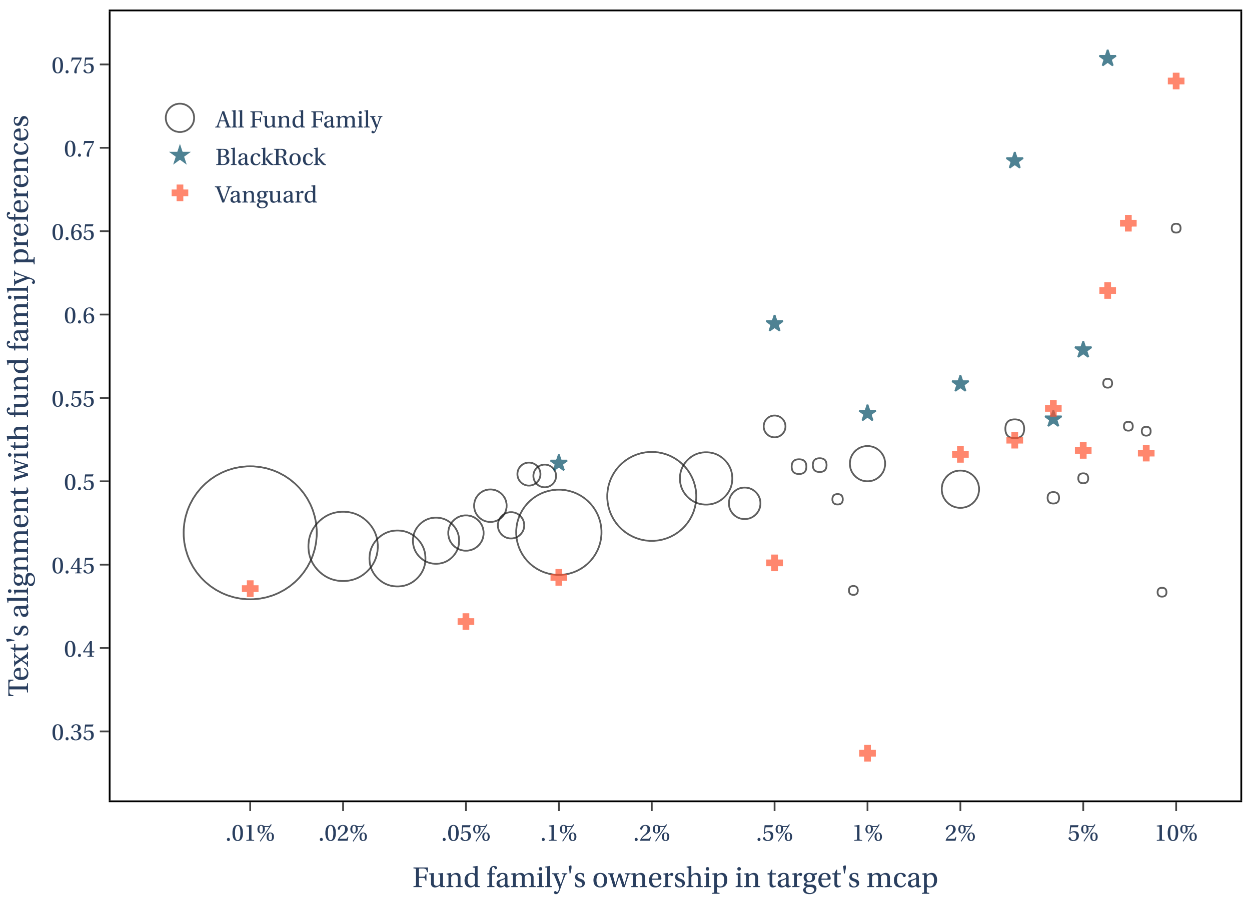
Table 4 summarizes the sample set. The average holdings by fund family, which are invested, is 0.49% of the target's market cap; the standard deviation being 1.37 percentage points. The median holdings is 0.06%, and the 75th percentile is at 0.26%, indicating that the data set is left-skewed bounded by zero. Out of the 12,582 fund family holdings observations, 11,113 or 88% are less than 1%. Putnam Investments has the maximum holdings at 18.58% in Altisource Residential Corporation, when Oliver Press Partners attacked Altisource in January 2016. In the full sample, 54,790 observations have zero holdings. The mean is 0.09%, with a standard deviation of 0.63 percentage points. The two samples are comparable in terms of

the attack text's alignment with fund family preferences. On average, the attack text is aligned 48% with the fund family preferences i.e., the activist will garner the support of 48 out of 100 mutual fund portfolios in a fund family during the campaign. The median number is similar, at 44%. The 25th percentile is 7%, and the 75th percentile is 94%. The text's alignment bunching at zero and one is in line with how most portfolios in a fund family vote as a block (Cai and Walkling 2011; Rothberg and Lilien 2006).

Going ahead, I focus on the dataset containing all the fund families. The full dataset takes care of the survivorship bias present in the subsample of fund families with a stake in the target. I hypothesize that activists care more about the preferences of fund families that own more shares. An activist would have their attack text more aligned to a fund family with a 5% stake than a fund family with a 0.1% stake. However, what if the text is more aligned to the preferences of fund families that are not even invested? To alleviate survivorship bias concerns, I focus on the full sample of data. Separately, Appendix D reports results for the subsample of fund families with a stake in the target.

Figure 4 plots the average estimated attack text's alignment text for different fund family holdings. The bulk of the dataset is in the left corner, with only 279 (or 2.19%) observations having a fund family that owns more than 5% of target shares. To flesh out numbers close to zero, I plot the x-axis in logarithmic terms. Across holdings, the average alignment is above 40%. The plot has an increasing trend, indicating a positive association between fund families owning shares in the target and attack text pandering to families' preferences. For the subsample of BlackRock and Vanguard, the increasing trend in the attack text's alignment when holdings increase persists. I get similar relationships for Charles Schwab, Fidelity, and State Street (not included in the figure owing to space constraints). The SVR method predicts, on average, BlackRock, compared to Vanguard, is more likely to support activists. BlackRock's higher activist support is in line with how it had voted more aggressively against management recommendations in shareholder proposals, illustrated in Table 1. The figure suggests that activists talk about issues important to the fund families that hold more shares.

Figure 4:
The attack text's alignment with fund family preferences is positively associated with the fund family's holdings in the target.
The figure plots the average attack text's alignment with fund family preferences for holdings between 0 to 10%. The alignment is based on the fund family's voting patterns on shareholder proposals in the two years before the attack. Holding represents the fund family's ownership in the target stock as a percentage of the target's market cap. Holdings are rounded to the nearest tick mark, and the corresponding alignments are averaged. The radius for all fund family series corresponds to the number of observations around the tick.



4 Evidence of campaigns aligned with larger shareholders

4.1 Activists focus on investors that hold more of target shares

Theoretically, pandering to larger shareholders allows the activists to consider the preferences of a few fund families, and yet cover a big chunk of the voting bloc. Alchian and Demsetz (1972) specify that the temporary coalescence of share votes into voting blocs is required to displace the existing management or modify managerial policies. Tom Ball, CEO of Vanderbilt Consulting, notes in an interview: "With the increasing concentration of ownership it is the top ten holders who will win it or lose it for you ... support on one side or the other of the largest two or three out of the top five will make the difference" (TheStreet 2017). Focusing on a few institutions also reduces the coordination costs of activism (e.g., during the proxy solicitation process). However, focusing on the larger fund families could also mean a wasted effort. The larger fund families often have a conflict of interest. In particular, a fear of losing the business of corporate pension plans, one of the largest investors in index funds, may deter such institutions from supporting activists (Ashraf, Jayaraman, and Ryan 2012; Cvijanović, Dasgupta, and Zachariadis 2016; Davis and Kim 2007). These institutions also suffer from a lack of incentive to affect change and are hesitant to expend additional resources required for engaging with the activists (Lund 2017). Moreover, despite the SEC ban on companies to dole out information selectively, fund managers might still get a cold shoulder from executives if they are too critical. A significant number of larger institutional investors are also skeptical of activists' demands, including requests for increased debt and payouts, which they see as shortsighted.

I find that activists tailor their communications to the preferences of fund families that own more shares in the target. Specifically, I estimate the following:

$$\widehat{Align}_{a,f} = \beta Holding_{a,f} + \delta_a + \delta_f + \epsilon_{a,f} \tag{4}$$

where $\widehat{Align}_{a,f}$ is the predicted alignment of the attack text, a, with the fund family's, f, pref-

erences. *Holding* refers to the fraction of the target's market cap the fund family owns before an attack. δ_a and δ_f are attack level and fund family level fixed effects. Because both *Align* and *Holding* could be correlated across observations of a particular attack and because the estimation errors, ϵ, might exhibit serial correlation, I cluster the standard errors at the attack level. However, subsequent findings are robust to not clustering or clustering at other levels (e.g., fund family).

Our main identification concern is that of omitted variables. If the fund family's holding in target is correlated with activist-, target-, or fund family-level characteristics that affect attack text's alignment with the family (or how closely the attack speaks about issues important to the fund family), then my estimate of interest, β could reflect these omitted variables rather than an effect of holdings on campaign tailoring. For example, if activists align their attacks with major fund families in terms of total net asset, such as BlackRock, Vanguard, etc., and these families are often the larger shareholder in an average attack, then *Holding* and *Align* could be positively correlated for this reason rather than because the activists actively tried to align the attack with the larger shareholders in target.

The inclusion of attack and fund family fixed effects, however, allows me to control for a number of these potential omitted factors. The attack fixed effects control for any activists characteristics that could affect the attack's tailoring to align with fund family preferences, including activist's skills and support they get from ISS. The attack fixed effects also control for any characteristics of the target (e.g., target's, performance, strategy, ownership structure, etc.) at the time of the attack that might matter for how institutions feel about certain phrases related to the target. The fund family fixed effects control for any differences in a fund family's overall likelihood of voting against the management, which can vary considerably across institutions (Brav, Jiang, Li, and Pinnington 2018; Kedia, Starks, and Wang 2020). Hence, by including both fixed effects, the coefficient of interest, β will only be identified using variation in how attack text's alignment for a given attack varied as a function of each fund family's level of holdings and variation in how alignment by a fund family varied as a function of its

Table 5:
Activists align their communications to the preferences of larger shareholders
This table reports estimates of regression of the attack text's alignment with fund family preferences on the fund family's holdings in targets. Specifically, I estimate:

$$\widehat{Align}_{a,f} = \beta Holding_{a,f} + \delta_a + \delta_f + \epsilon_{f,a}$$

where $\widehat{Align}_{a,f}$ is the predicted alignment of the attack text, *a*, with the fund family, *f*, preferences. *Holding* is the percent of equity the fund family owns of the target before the attack, obtained from the CRSP database. δ_a and δ_f represent attack level and fund family level fixed effects, respectively. The sample consists of attacks, identified using attack filings, over the 2004–2019 period. Corresponding fund families include all the families that have voted in at least a hundred shareholder proposals in the two years prior to the attack. The independent variable is scaled by the standard deviation of the underlying variable, meaning the coefficient can be interpreted as the effect of a one-standard-deviation change in the determinant. Standard errors, $\epsilon_{a,f}$, are clustered at the attack level, and t-statistics are reported in brackets below the coefficient estimates. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

	Attack text's alignment with fund family preferences			
	(1)	(2)	(3)	(4)
Fraction of target mcap	0.0059***	0.0042**	0.0073***	0.0047**
held by fund family	[3.81]	[2.19]	[3.15]	[2.48]
Attack FE		Yes		Yes
Fund family FE			Yes	Yes
Observation	66,432	66,432	66,432	66,432
R^2	0	0.135	0.094	0.224

holdings in targeted firms' stocks.

Table 5 reports estimates for the association of the predicted attack text's alignment on fund family holdings using variants of Equation 4. The percentage of shares held by fund families in the target is significantly (at 1% level) associated with the attack text's alignment with the family. A one-standard-deviation increase in fund family holdings in targeted shares, which is approximately 0.63 percentage points relative to an average of 0.09%, is associated with around a 0.47 percentage point increase in the attack text's alignment with the fund family preferences. Relative to the average alignment of 48%, this corresponds to a sizable increase. The coefficient for holdings is positive and significant at either level of fixed effects.

Separately, Appendix Table 13 reports similar results for a sub-sample dataset, which includes only fund families invested in the target.

An alternative scenario that could drive the positive association is if the larger fund families such as BlackRock, Fidelity, Vanguard, etc. are, in general, more likely to agree with the activists, or equivalently have a higher alignment with the attack text. Section 2.2 reports otherwise. The fund families with higher holdings, which often happen to be passive index investors, are less likely to vote against the management in shareholder proposals. As the training data is not skewed, the pattern is unlikely to occur in the predicted or test sample. In Table 5, the regression results with the fund family fixed effect (3) and (4) further support the hypothesis. For a particular fund family, the attack text's alignment increases by 1.1 percentage points (1.59×0.724) for every percent increase (1%/0.63% = 1.59 standard deviation) in the family's holdings in the target. Thus, the text is geared towards the preferences of major institutional holders in the target, instead of just the major institutional investors in general. Appendix C lists examples of activists using phrases to appeal selectively to BlackRock, Fidelity, and Vanguard when the fund families own a larger stake in the target.

4.2 Activists learn from interactions with fund families

Despite the publicly available voting and soft information activists possess about fund family preferences, strategically aligning attack texts to these preferences is not easy. First, there is a limit to how aligned an attack text could get for a particular fund family. Moreover, variations in preferences across fund families make it harder for the activist to align the text well with all the stakeholders. Nonetheless, there are benefits of strategically aligning the text, including coalescing of large shareholders' votes. As such, the hedge fund institutions are likely to mimic the successful strategy (Alchian 1950), and with experience, hone the skill. So, is there a learning curve where activists use this campaign tailoring strategy better as they mature?

To test the activist's learning curve, I define $NumInteraction_{a,f}$, which indicates the interaction count for an activist with a fund family, f, before a particular attack, a. I start by sort-

ing the attacks in terms of attack date and assigning $NumInteraction_{a,f}$ equal to zero. For each attack by the activist, if the fund family owns more than a percent of shares in the target, $NumInteraction_{a,f}$ increases by one. In essence, *NumInteraction* measures the number of times an activist has interacted with a fund family when the fund family owns significant shares in the target. I employ:

$$\widehat{Align}_{a,f} = \beta NumInteraction_{a,f} + \delta_a + \delta_f + \epsilon_a \tag{5}$$

where $\widehat{Align}_{a,f}$ is the predicted alignment of attack text, *a*, with the fund family, *f*, preferences. *NumInteraction* is the number of times the fund family has been a significant shareholder in the activist's attacked targets. δ_a and δ_f are attack level and fund family level fixed effects. Finally, I adjust the standard errors, $\epsilon_{a,f}$, for clustering at the attack level.

Table 6 demonstrates that the activists align their texts better when they interact more with a fund family. For every interaction between an activist and a fund family owning more than a percent of target shares, the attack text's alignment with the fund family increases by 0.9 percentage point. The increase is substantial, compared to the average attack text's alignment of 48 percentage points. The results are robust to fixing attack and fund family level variations. Thus, the activists are more willing to include phrases that appeal to fund families with whom they have had an attack relevant interaction before.

The results support Appel, Gormley, and Keim (2019), who point out that activists have learned through their repeated interactions, and are able to tailor their campaign tactics and goals to reflect priorities of long-term investors. Howard Sherman, CEO of Institutional Shareholder Services, also agrees that "these hedge funds are looking for returns, the push for governance is coming from a larger and larger number of public pension funds and investment managers" (InstitutionalInvestor 2006). The shift in tactics could explain the increased success of activists and the increased openness of some institutions to activists' demands. For example, in the 2015 letter to corporates, Larry Fink, CEO of BlackRock, stressed that short-term thinking is getting in the way of long-term business growth. (BlackRock 2015). In contrast,

Table 6:
Activists learn from interactions with fund families to tailor their communications better
This table reports estimates of regression of an attack text's alignment with the fund family preferences on the number of times the activist has interacted with a fund family. Specifically, I estimate:

$$\widehat{Align}_{a,f} = \beta NumInteraction_{a,f} + \delta_a + \delta_f + \epsilon_{a,f}$$

where $\widehat{Align}_{a,f}$ is the predicted alignment of attack text, *a*, with the fund family, *f*, preferences. *NumInteraction* is the number of times the fund family owned more than a percent of target shares in an attack initiated by the activists. The fund family ownership data is obtained from the CRSP database. δ_a, and δ_f are the attack level and fund family level fixed effects. The sample consists of attacks, identified using attack filings, and corresponding fund families over the 2004–2019 period. Corresponding fund families include all the families that have voted in at least a hundred shareholder proposals in the two years prior to the attack. Standard errors, $\epsilon_{a,f}$, are clustered at the attack level, and t-statistics are reported in brackets below the coefficient estimates. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

	Attack text's alignment with fund family preferences		
	(1)	(2)	(3)
Number of Interaction	0.0098***	0.0084**	0.0088**
	[4.89]	[2.27]	[2.15]
Attack FE		Yes	Yes
Fund family FE			Yes
Observation	66,432	66,432	66,432
R^2	0	0.135	0.224

Fink admitted in 2018 that the interactions between targets and activists are often productive for long-term investors like his funds (Reuters 2018).

5 Impact of campaign tailoring

5.1 Funds pay attention when activists speak to their preference

The impact of campaign tailoring on the fund families' willingness to engage with activism is ambiguous. There is evidence of human predisposition to select claims adhering to their system of beliefs and ignore dissenting information (Bessi et al. 2015; Del Vicario et al. 2017,

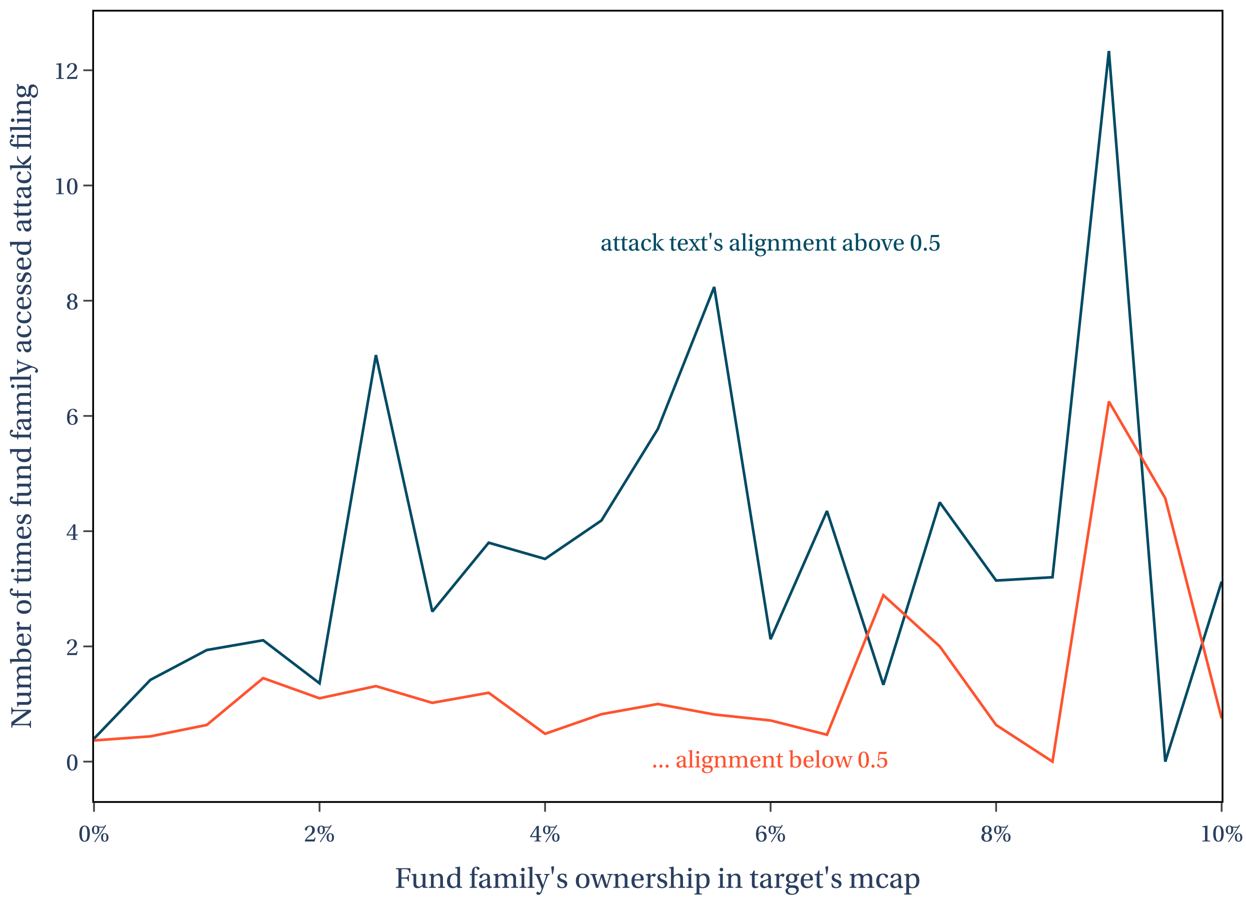
2016). In contrast, the growing popularity of passive investment strategies and institutions' increasingly diverse holdings raises questions about how actively institutions monitor individual attacks. Even the largest fund families have only 20 or so people in their governance departments, or around one per thousand invested companies, to determine how proxies should be voted (FT 2016). Fund families' diverse investments make it harder for activists to engage the families, even when activists raise issues tailored to families' preferences.

I use fund families' attack filings access on the SEC's EDGAR server as a proxy for their attention on an attack. The SEC's Division of Economic and Risk Analysis (DERA) assembles information on internet search traffic for EDGAR filings through SEC.gov, covering the period February 14th, 2003, through June 30th, 2017. I use a linking table from Digital Element to assign IP addresses in the log files to fund families. The linking table contains organizations' names and registered IP addresses as of December 31st, 2016. I follow Iliev, Kalodimos, and Lowry (2020) to identify EDGAR activity related to governance research. Using the accession number included in attack filings, I make a list of attack documents. I sum the number of times a fund family viewed one of the attack documents from the date the attack begins to 30 days after the attack ends. The attack beginning and end date are defined as the first and last date of the attack filing date.

In total, I gather fund views of attack documents for 427 attacks, involving 115 unique fund families. For each attack, I include fund families that have (i) voted in at least a hundred shareholder proposals in the two years prior to the attack and (ii) checked filings of at least 1% of their investments. My data set contains 244 thousand attack-fund family pairs, with an average of 1.04 views by a fund family for an attack. Counting only the positive views, I have 40 thousand attack × fund family data points, involving 73 unique fund families on 278 attacks. The average positive number of views for a fund family is 6.39 views per attack. Appendix A.3 describes, in more detail, the procedure to extract attack filings access by the fund family on the SEC server.

Figure 5 illustrates differences in the fund families' attention, based on the attack text's

Figure 5:
Fund families conduct more research about attacks that are well-aligned.
This figure plots the number of times fund families accessed attack filings on the SEC.gov server, averaged at each half percentage point holdings. The data for fund families access of SEC filings is available from DERA. The period considered for each attack spans the date the attack begins to 30 days after the attack ends. The attack's beginning (end) date is the first (last) date of attack filing by the activist. The attack text'alignment with fund family preferences is based on the family's voting choices on shareholder proposals.



alignment. When an attack is more aligned, indicated by an alignment score above 0.5, the fund family is more likely to access attack filings on SEC.gov. The result is consistent across different fund families' holdings in the target. For similar investments, fund families are more likely to pay attention when the attack text speaks to their concerns. To formally test whether a higher alignment is associated with more attention, I estimate:

$$View_{a,f} = \beta \widehat{Align}_{a,f} + \delta_a + \delta_f + \epsilon_{a,f} \tag{6}$$

where $View_{a,f}$ represents the number of times attack documents, a, were accessed by a fund family, f. $\widehat{Align}_{a,f}$ is the predicted alignment of attack text with the fund family preferences. δ_a and δ_f are attack level and fund family level fixed effects. Finally, I adjust the standard errors, $\epsilon_{a,f}$, for clustering at the attack level. The model is estimated over the January 2004–June 2017 period.

Table 7 reports that fund families pay more attention to the attacks that cater to their preferences. The estimated coefficient is positive and statistically significant (at the 1% level) using either attack level variation or fund family level variation. Specifically, a one-standard-deviation increase in the predicted alignment of attack text with fund family preferences, which is approximately 40 percentage points relative to an average of 48%, is associated with 0.11 more views of attack filings on the SEC.gov website. Considering that the average number of views for an attack is 0.48, the increase represents a 23% higher EDGAR activity by the fund family. The increased attention has implications for the attack, as fund families are usually nonchalant about activism. For example, BlackRock, Fidelity, and Vanguard decided to forgo GameStop ballots by keeping shares on loan (WSJ 2020b).

Gormley and Jha (2020); Iliev, Kalodimos, and Lowry (2020); Malenko and Shen (2016) demonstrate that fund families pay more attention when they have a higher stake in a firm. Moreover, Section 4.1 shows that the alignment is dependent on the equity share the fund owns. To mitigate the omitted variable problem, I include the fund families' equity share in the target as a control in part (4), (5), and (6) of Table 7. Even for similar investments in the tar-

Table 7:
Fund families conduct more research on attacks tailored to their preferences
This table reports estimates of regression of fund family access of attack text filings on the attack text's alignment with fund family preferences. Specifically, I estimate:

$$View_{a,f} = \beta \widehat{Align}_{a,f} + \delta_a + \delta_f + \epsilon_{a,f}$$

where $View_{a,f}$ is the number of times a fund family, f, accessed attack filings, a, between the date the attack begins to 30 days after the attack ends. The attack's beginning (end) date is based on the first (last) date of attack filing by the activist. $\widehat{Align}_{a,f}$ is the attack text's alignment with fund family preferences. δ_a and δ_f are attack level and are fund family level fixed effects. The sample consists of attacks, identified using SEC filings, and corresponding fund families over the 2004–2019 period. Data for fund families' access of filings on SEC.gov is available via DERA. Columns (4), (5), and (6) control for fund family holdings in the target. Independent variables are scaled by the standard deviation of the underlying variable, meaning coefficients can be interpreted as the effects of a one-standard-deviation change in the determinant. Standard errors, $\epsilon_{a,f}$, are clustered at the attack level, and t-statistics are reported in brackets below the coefficient estimates. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

	Number of times fund family viewed attack filings on SEC.gov					
	(1)	(2)	(3)	(4)	(5)	(6)
Attack text's	0.0856***	0.0849***	0.115***	0.0774***	0.0771***	0.112***
alignment	[4.24]	[3.59]	[4.69]	[3.86]	[3.36]	[4.57]
Fraction of target mcap				0.381***	0.365***	0.251***
held by fund family				[18.98]	[7.01]	[5.25]
Attack FE		Yes	Yes		Yes	Yes
Fund family FE			Yes			Yes
Observation	34,174	34,174	34,173	34,174	34,174	34,173
R^2	0.001	0.106	0.163	0.011	0.115	0.167

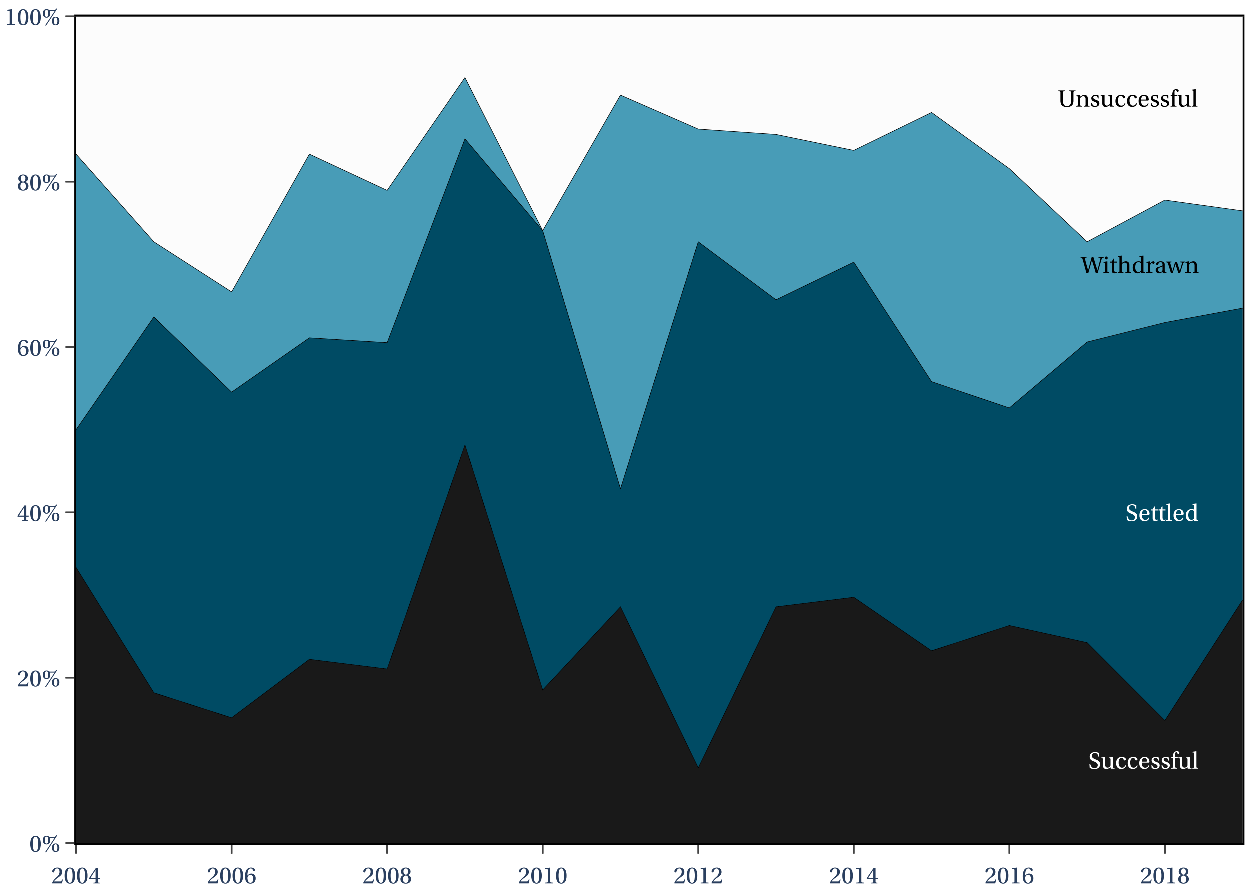
get, fund families to which an attack text is more aligned pay more attention to proxy filings. Appendix Table 14 reports the result of the analysis of the smaller sub-sample of invested fund families.

5.2 Fund families support attacks geared towards them

A priori, the impact of campaign tailoring on actual fund family voting is not clear. First, in psychology, taste and trust are considered strongly related, i.e., people are more likely to accept familiar information as true (Jøsang, Quattrociocchi, and Karabeg 2011; Swire, Ecker, and Lewandowsky 2017). Therefore, a fund family is likely to see an attack as credible if the activist discusses issues important to the fund family. The opposite is also plausible. The media scrutinizes mutual funds' voting decisions if the attack proposals pertain to issues that fund managers have raised in interviews. For example, BlackRock voting decisions on climate proposals, and State Street's stance on gender diversity proposals, are often covered more in the media than vice versa. As such, activist support from fund managers, especially the active managers, could draw attention to investments that haven't fared well, while the managers held on to these stocks for years. The negative media attention for one of their investments can nudge the fund family not to be involved in activism and stick to the default of supporting the target's management.

To analyze whether tailoring an attack's text affects fund families' voting decisions, I restrict my analysis to attacks that reached the voting stage. I use CapitalIQ available from S&P to gather information on how the attacks panned out. I get results for 461 attacks in my sample. The CapitalIQ platform classifies attack results into four categories: (i) Successful: the activist's proposals win the shareholder election. (ii) Settled: targets and activists discuss and compromise, without going into a formal election. Settlement often occurs when the target feels that the activist's case is strong and tries to avoid the embarrassment of losing the election. (iii) Withdrawn: activists get the signal of insufficient support and cut losses by withdrawing the case. (iv) Unsuccessful: the activists participate in the election and are unable to

Figure 6:
Distribution of attack outcomes remains persistent.
The stacked area plots the outcome of attacks over the 2004–2019 period. Attacks are assigned to the year when they began, i.e., the earliest date of SEC filings pertaining to the attacks. The information on attacks' outcomes is collected from S&P CapitalIQ.



secure required votes.

Figure 6 shows the distribution of attack outcomes. 63% of the attacks are either successful or settled over the 2004–2019 period, the two categories peak at 85% for attacks beginning in 2009. The number exceeds 60% throughout, except for attacks that began in 2011. Only about half of the attacks go for proxy voting. Gantchev (2012) estimates a campaign ending in a confrontational proxy fight costs $10.71 million on an average. Moreover, targets fear reputation costs, which could be high for election losses. Therefore, both the parties, activists and targets, try to settle out of a proxy contest. Over the 2004–2019 period, 199 attacks went to actual shareholder voting, marked Successful and Unsuccessful. I use the voting information

on these attacks to compare text-based voting predictions with fund families' actual voting choices.

To identify the proposals associated with an attack, I start with all the shareholder meetings for the target. I filter out meetings that have (i) no shareholder sponsored proposals, (ii) occurred more than 30 days before the attack's end date, or (iii) occurred more than 365 days after the attack's end date. I use a 30-day window because activists often communicate with shareholders after the voting to inform them of meeting results and offer gratitude. After filtering, I choose the first meeting after the attack's beginning date. I select all the shareholder proposals that do not contain the string "Management Nominee" in the ISS voting database. I aggregate the activist support at the attack level, i.e., if the fund family supported one out of three activist proposals, the *SupAct* for the attack will be 0.33. I get a total of 1,457 voting records of fund family voting on attack proposals. To test my hypothesis that fund families vote more favorably on attacks which speak to their preferences, I estimate:

$$SupAct_{a,f} = \beta \widehat{Align}_{a,f} + \delta_a + \delta_f + \epsilon_{a,f} \tag{7}$$

where $SupAct_{a,f}$ is the fraction of mutual funds, for a fund family, *f*, that supported activists' proposals in the shareholder meeting following an attack, *a*. $\widehat{Align}_{a,f}$ is the attack text's alignment with fund family preferences. δ_a and δ_f are attack level and fund family level fixed effects. Finally, I adjust the standard errors, $\epsilon_{a,f}$, for clustering at the attack level.

Results, reported in Table 8, indicate that the attack text's alignment with the the fund family is positively associated with the actual voting outcome. The estimated coefficients are positive and statistically significant (at the 1% level) for attack and fund family level fixed effect. Among the fund families who voted in attacks, a one-standard-deviation increase in the attack's alignment with fund family preferences, which corresponds to 44 percentage points relative to the average 56%, is associated with a three percentage point increase in the fund family's actual support for the activist, relative to the average 53%. Thus, the text-based measure of voting outcome is in line with the actual voting outcome, robust to within attack or

Table 8:
Fund families support attacks geared to their preferences
This table reports estimates of regression of fund family activist support on the attack text's alignment with fund family preferences. Specifically, I estimate:

$$SupAct_{a,f} = \beta \widehat{Align}_{a,f} + \delta_a + \delta_f + \epsilon_{a,f}$$

where $SupAct_{a,f}$ is the fraction of mutual funds, for a fund family, *f*, that supported activists' proposals in the shareholder meeting following an attack, *a*. Attack proposals include shareholder proposals that were part of shareholder meeting after the attack. $\widehat{Align}_{a,f}$ is the attack text's alignment with fund family preferences. δ_a and δ_f are attack level and fund family fixed effects. The sample contains observations of fund families voting on attack proposals over the 2004–2019 period. I use SEC filings to get all the attacks, and CapitalIQ to identify the ones that went to a voting stage. The independent variable is scaled by the standard deviation of the underlying variable, meaning the coefficient can be interpreted as the effect of a one-standard-deviation change in the determinant. Standard errors, $\epsilon_{a,f}$, are clustered at the attack level, and *t*-statistics are reported in brackets below the coefficient estimates. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

	Actual activist support		
	(1)	(2)	(3)
Attack text's alignment	0.0186*	0.0468***	0.0310***
	[1.88]	[4.44]	[3.2]
Attack FE		Yes	Yes
Fund family FE			Yes
Observation	1457	1453	1419
R^2	0.002	0.557	0.611

fund level variations. The findings support Kamenica and Gentzkow (2011); McCloskey and Klamer (1995), who argue that persuasion plays a key role in voting decisions. A higher movement required for the text's alignment to change the fund family's actual support indicates that other factors, apart from attack text content, dictate fund family voting. Other factors that impact voting decisions and limitations associated with a text-based measure are discussed in Appendix E.

5.3 Attacks geared to larger investors succeed

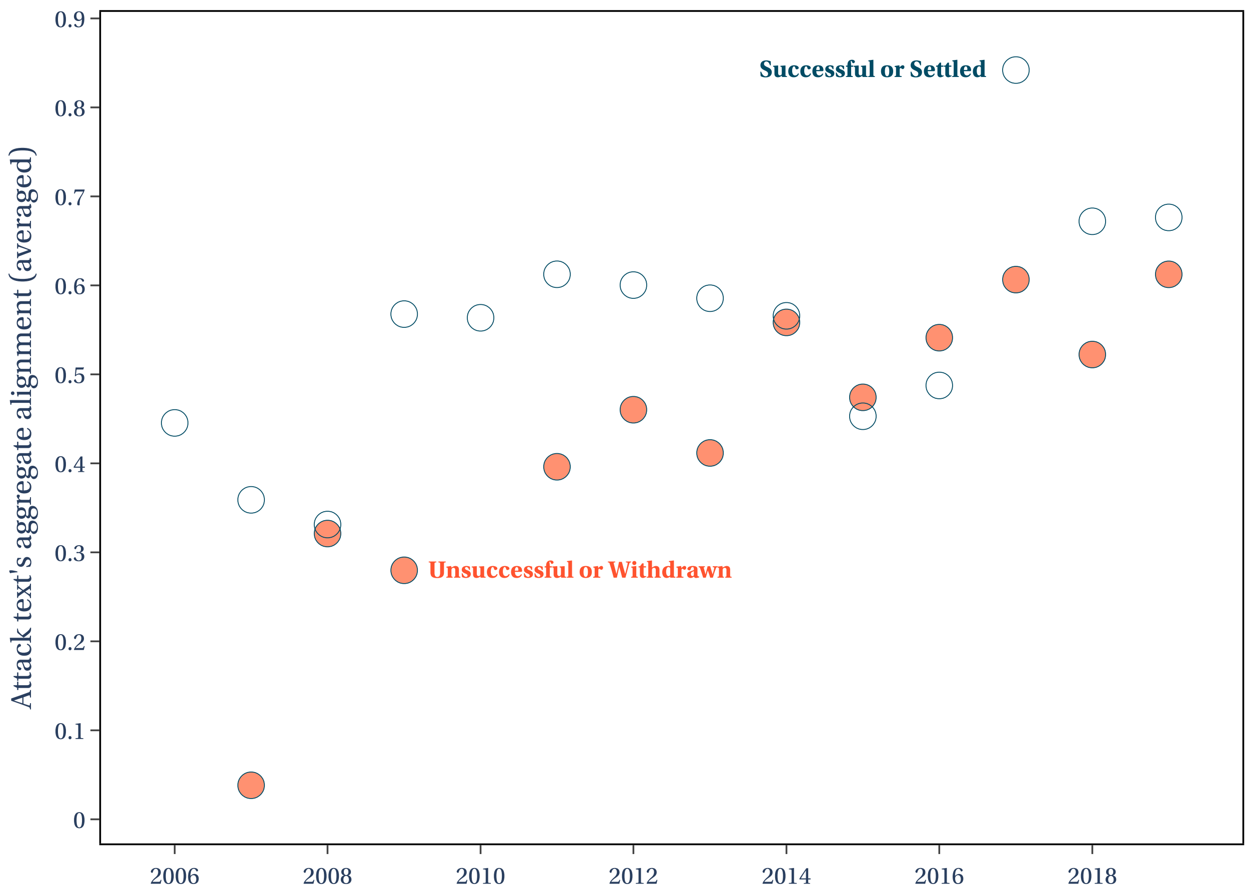
Section 5.2 shows that the persuasion catches fund families' votes. A nod from a major shareholder in the target could tip the attack in the activist's favor. Moreover, gaining support from a major shareholder adds credibility to the campaign, persuading other shareholders to commit. In 2019, EQT's largest shareholder, T. Rowe Price, issued a press release stating support for dissident Rice Group nominees. A week later, shareholders elected all seven Rice nominated directors at EQT's annual meeting. The question begets: does the pandering help activists win? To answer the question, I start with a simple measure of how well an activist aligns the attack text with fund family preferences. I define:

$$AgAlign_a = \sum_f \widehat{Align_{a,f}} \times \frac{Holding_{a,f}}{\sum_f Holding_{a,f}} \tag{8}$$

where *AgAlign* is aggregate align for an attack, *a*, based on the attack's alignment with a fund family weighted by the fund family's holdings. It measures aggregate mutual fund support, i.e., what fraction of the mutual fund's vote will the activist gather, based on the attack text. An attack that is well-aligned with fund families owning larger shares in the target will have a higher *AgAlign*.

A drawback of *AgAlign*, as a parameter for attack outcome, is that it is volatile for attacks where the mutual fund presence is not significant. For an attack where mutual funds own less than five percent of market cap, even when the attack is well-aligned with *AgAlign* close to one, the small ownership is usually not sufficient to have an attack level impact. Brav, Dasgupta, and Mathews (2019) also show in their model that campaigns succeed if the measure of shares that engage is above a threshold. To circumvent the issue, I interact *AgAlign* with a dummy for ownership in the target. The ownership dummy, *OwnDum*, is one if the mutual funds, whose alignments aggregate to *AgAlign*, own more than the sample average, 14.3%, in target shares. Out of the 419 attacks that had a fund family with voting records as a shareholder, 195 have ownership dummy of one. Appendix Figure 12 shows that the result holds for changing the

Figure 7:
Attacks that are aligned well with the larger shareholders are more likely to win.
This figure plots the attack text's aggregate alignment, averaged for each year based on the attack's outcome. The sample includes attacks with at least 14.3% (the mean holding for full sample) of target shares held by one of the fund families with available voting records. The aggregate alignment is the attack text's alignment weighted by fund family holdings, defined in Equation 8. The shade of the bubble represents the outcome of the attack. In 2006 and 2010, no attacks above the cutoff holding were unsuccessful or withdrawn.



cutoff of ownership dummy.

Figure 7 is a scatter plot for aggregate alignment, averaged each year based on the attack's outcome. I divide the attack outcomes, shown in Figure 6, into two groups: Settled/Successful and Withdrawn/Unsuccessful, as the outcomes within the groups are considered equivalent in activism literature (Brav, Jiang, and Kim 2015b). Attacks geared more towards the larger fund families' preferences, identified by a higher aggregate alignment for the attack text, succeed more often. Every year, except for 2015, the average aggregate alignment is higher for attacks that succeed. For the sub-sample of twelve attacks, where an activist attacked the same target twice with differing outcomes, the average aggregate alignment for Successful/Settled attacks is 56%, compared with Unsuccessful/Withdrawn attacks at 39%. I formally test whether the attack text's alignment is associated with activist success by employing:

$$Win_a = \gamma AgAlign_a + \lambda OwnDum_a + \beta AgAlign_a \times OwnDum_a + \epsilon_a \tag{9}$$

where Win_a is a dummy equal to one if the activist wins the attack, a, i.e., Win_a is one if the attack is Successful or Settled. *AgAlign* is the aggregate mutual fund support for the activist, based on the attack text. *OwnDum* denotes the ownership dummy, which is one if the mutual funds, whose alignment aggregate to *AgAlign*, own more than the sample average, 14.3%, in target shares. By including the additional interaction, the coefficient on *AgAlign* will now capture the importance of text alignment for all the other attacks. In contrast, the sum of the coefficients on *AgAlign*, *OwnDum*, *AgAlign x OwnDum* will capture the importance of text alignment for the attacks with above-average mutual fund holdings. The standard errors, ϵ_a, are robust and computed with the sandwich estimator of variance.

Table 9 shows that attacks, which are more in line with larger fund families' preferences and have sufficient mutual fund holdings, are indeed more likely to succeed, shown in Column (2). Specifically, among attacks for which the mutual fund ownership is at least the average, a one-standard-deviation increase in *AgAlign*, or 28 percentage points, is associated with a 9.4 (11.1 + 0.48 - 2.18) percentage point increase in the likelihood of the activist winning the proxy

Table 9:
Activists are more likely to win when they pander to larger shareholders
This table reports estimates of regression of attack outcomes on attack text's alignment weighted by the fund family holdings. Specifically, I estimate:

$$Win_a = \gamma AgAlign_a + \lambda OwnDum_a + \beta AgAlign_a \times OwnDum_a + \epsilon_a$$

where Win_a represents a dummy, which is one if the result of the attack, *a*, is Successful or Settled. *AgAlign* is the holdings-weighted attack text's alignment with fund families. *OwnDum* is the ownership dummy, which is one if the mutual funds with voting information own more than the average, 14.3%, of target shares. The sample consists of all the attacks, identified using SEC filings, that went to a voting stage over the 2004–2019 period. I use CapitalIQ to gather outcomes. The independent variable, *AgAlign*, is scaled by the standard deviation of the underlying variable, meaning the coefficient can be interpreted as the effect of a one-standard-deviation change in the determinant. The standard errors, ϵ_a, is robust and computed with the sandwich estimator of variance. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

	Indicator for activist win	
	(1)	(2)
Aggregate alignment	0.0214	−0.0218
	[0.90]	[-0.71]
Ownership dummy		0.0048
		[0.10]
Aggregate alignment × Ownership dummy		0.111**
		[2.25]
Observation	419	419
R^2	0.002	0.014

attack. For reference, the average *AgAlign* is 53% for the sample, i.e., for an average attack text, the activist could expect 53% of all the votes cast by mutual funds. The 9.4% increase is significant given the average *Win*, or an attack's likelihood of being successful or settled is 63%. The coefficient for ownership dummy is not significant, indicating that attack outcomes are not significantly different across the *OwnDum* cutoff. The relationship also does not hold for attacks that do not have mutual fund investments above the 14.3% threshold, illustrated in Column (1).

6 Validation and robustness

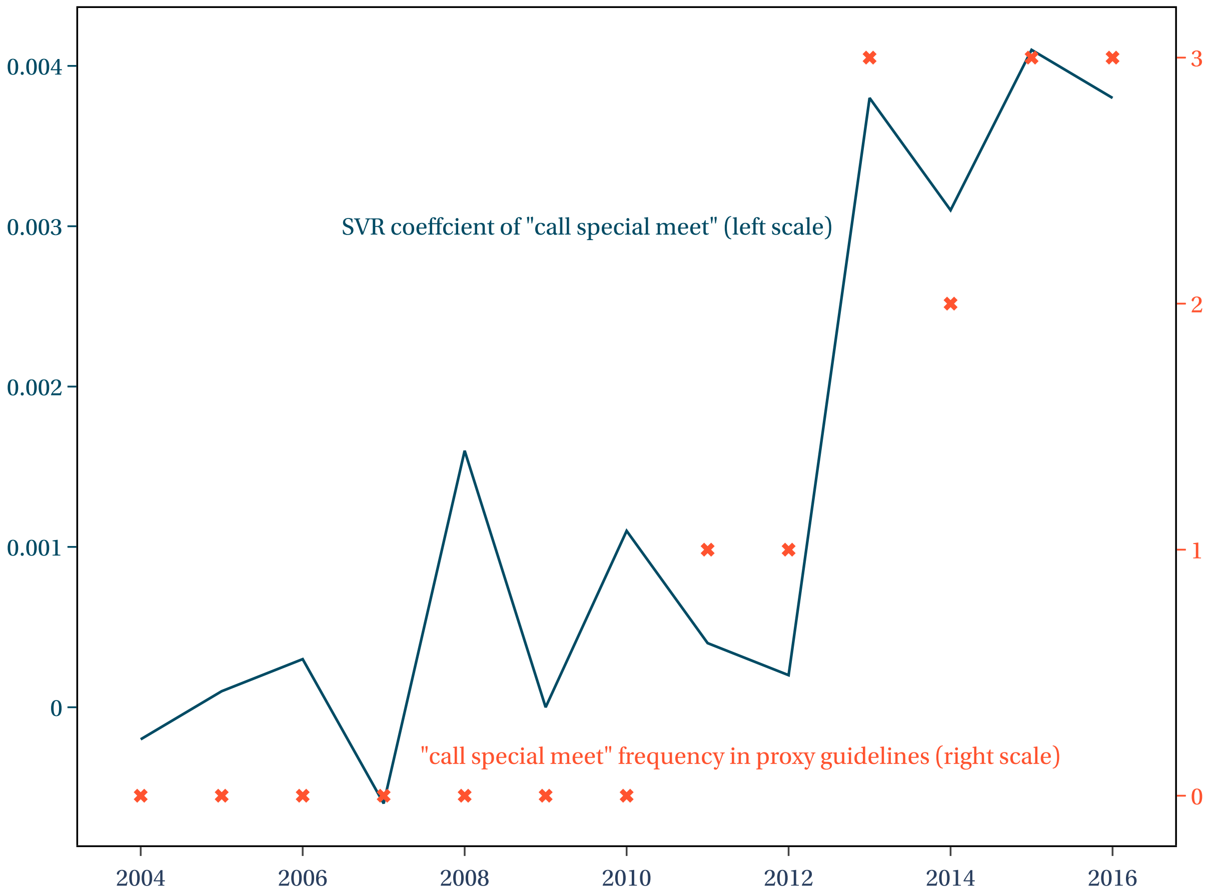
6.1 SVR coefficients follow proxy voting guidelines

Mutual funds distribute funds' prospectus to shareholders yearly, describing, among other things - risks, investment strategies, and proxy voting guidelines. Funds file the prospectus with the SEC as a post-effective amendment form or 485BPOS. Proxy guidelines describe fund policies on different corporate governance issues such as director elections, auditor approvals, compensation issues, corporate structure, shareholder rights, social policy, etc. Proxy guidelines reveal variations in preferences for a fund family across years, along with variations across fund families.[11] Proxy guidelines across mutual funds within a fund family remain mostly consistent for a given year. Therefore, to gather voting policy text for a fund family, I look for the prospectus of the biggest mutual fund that is part of the fund family. I search for cues such as "Proxy Voting Guidelines," "Proxy Voting Policies and Procedures," etc., to extract the proxy voting guidelines. I get 378 proxy guidelines across 48 fund families over the 2004–2018 period.

It is unclear how the presence of specific phrases in the proxy guidelines should affect phrases' SVR coefficients. The occurrence of phrases such as "right to call shareholder meeting" indicates that the fund family wants to implement this right and would vote against the management if shareholder proposals contain this phrase. Higher against management voting would give these phrases a more positive coefficient. For example, Figure 8 shows variations in SVR coefficients of "call special meet" for Morgan Stanley across time. When Morgan Stanley mentions more about shareholders' right to have special meetings, the corresponding SVR coefficient is also higher. On the contrary, sometimes fund families also write about issues, such as climate, environmental, social, etc., which they feel are part of management de-

[11] For example, many fund families have become more inclined to vote against the management on social issues. Vanguard's willingness to oppose management on social issues has increased. Vanguard (2010): "regardless of our philosophical perspective on the issue, these decisions should be the province of company management unless they have a significant tangible impact on the value." Vanguard (2019): "The funds will evaluate each proposal on its merits and support those where we believe there is a logically demonstrable linkage between the specific proposal and long-term shareholder value of the company."

Figure 8:
SVR coefficients of “call special meet” follow Morgan Stanley’s proxy voting guidelines.
This figure plots the number of times the phrase “call special meet” is used in Morgan Stanley’s proxy voting guidelines and subsequent SVR coefficients. The SVR coefficients are calculated as of December 31st of the year after proxy guidelines are published. The coefficients are based on the fund family’s voting patterns on shareholder proposals in the two years before the calculation date.



cision prerogative and thus would vote with the management on those proposals. Therefore, mentions of these phrases in proxy guidelines could mean more negative SVR coefficients for the phrases.

To circumvent the ambiguity, I look for the absolute value of the coefficients. The rationale for this choice is that when a fund family mentions a particular phrase in their proxy guidelines, it is important to their voting decisions. As such, the fund family would vote more consistently when those phrases occur in a proposal. The consistency in voting assigns a higher absolute value to coefficients, more positive if the phrase is about supporting shareholders

and more negative if the phrase is about supporting management. In this section, I look for whether the SVR coefficients follow fund family policy guidelines by employing:

$$abs(\beta)_{p,f,t+1} = \beta Count_{p,f,t} + \delta_{f \times t} + \epsilon_{p,f,t} \tag{10}$$

where $abs(\beta)$ represents the absolute SVR coefficient associated with a phrase, *p*, for a fund family, *f*, at time t + 1. *Count* is the number of times a phrase appeared in the fund family's proxy guidelines text filed in year t. Since I use a two-year training period for SVR, I relate phrase counts from the proxy guidelines document to the SVR coefficients calculated at the end of next year. $\delta_{f \times t}$ shows fund family cross time level fixed effect and the errors, $\epsilon_{p,f,t}$, are clustered at the fund family level.

Table 10 reports that SVR coefficients are in line with their mentions in proxy voting guidelines. Within a fund family for a particular year, the coefficients are higher in absolute terms for phrases mentioned more in the proxy guidelines. For every mention of a phrase in the proxy guidelines, the absolute value of the coefficient increases by 0.007 percentage points. Some of the phrases that appear in the shareholder proposal may not appear in proxy guidelines of a fund family. One could argue that the phrases that do not appear could indeed be less significant for voting decisions, and that is why we have a positive coefficient associated with counts. To alleviate these concerns, in Columns (4)–(6), I include only those phrases which appeared at least once in the proxy guidelines text for the fund family. The results remain robust for the smaller sub-sample.

For reference, Fidelity mentions "golden parachute" five times in their 2017 proxy voting guidelines, compared to once by Vanguard in their 2017 guidelines. For an attack starting around December 2018, the SVR will assign 0.028 percentage points more text-predicted activist support for Fidelity, for every mention of "golden parachute" in the attack text. The key phrases are often included more than ten times in the communications. Furthermore, related phrases such as "poison pills," etc., co-occur with similar frequency and SVR coefficients. Therefore, an increase in mentions of a critical phrase in proxy guidelines is usually associated

Table 10:
SVR coefficients follow proxy voting guidelines
This table reports estimates of regression of absolute phrase coefficients on the number of times the phrase appeared in the fund family's proxy guidelines text. Specifically, I estimate:

$$abs(\beta)_{p,f,t+1} = \beta Count_{p,f,t} + \delta_{f\times t} + \epsilon_{p,f,t}$$

where $abs(\beta)$ represents the absolute SVR coefficient associated with a phrase, *p*, for a fund family, *f*, at the end of year t + 1. *Count* is the number of times the phrase appeared in the fund family's proxy guidelines text filed in year t. $\delta_{f\times t}$ represents fund family cross year fixed effect. (1), (2), and (3) show results for all the 9,832 phrases described in Section 3.1 for each fund family. The fund family sample is restricted to institutions that (i) have voted in at least a hundred shareholder proposals in the two years prior to the SVR calculation date, and (ii) have a proxy guidelines text available in 485BPOS filing. For (4), (5), and (6), I filter out phrases for a fund family if the phrase is not present in any of the fund family's proxy guidelines. The independent variable is scaled by the standard deviation of the underlying variable, meaning the coefficient can be interpreted as the effect of a one-standard-deviation change in the determinant. Standard errors, $\epsilon_{n,f,t}$, are clustered at the fund family level, and *t*-statistics are reported in brackets below the coefficient estimates. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

	Absolute SVR coefficient for the phrase × 10,000					
	(1)	(2)	(3)	(4)	(5)	(6)
Frequency of phrase	0.736***	0.734***	0.744***	0.477***	0.476***	0.477***
in proxy guidelines	[217.46]	[7.64]	[7.77]	[82.87]	[7.44]	[7.82]
Fund family FE		Yes			Yes	
Fund family × year FE			Yes			Yes
Exclude absent phrases				Yes	Yes	Yes
Observation	2,192,536	2,192,536	2,192,536	358,206	358,206	358,206
R^2	0.021	0.04	0.064	0.019	0.045	0.075

with up to a percent increase in text predicted activist support.

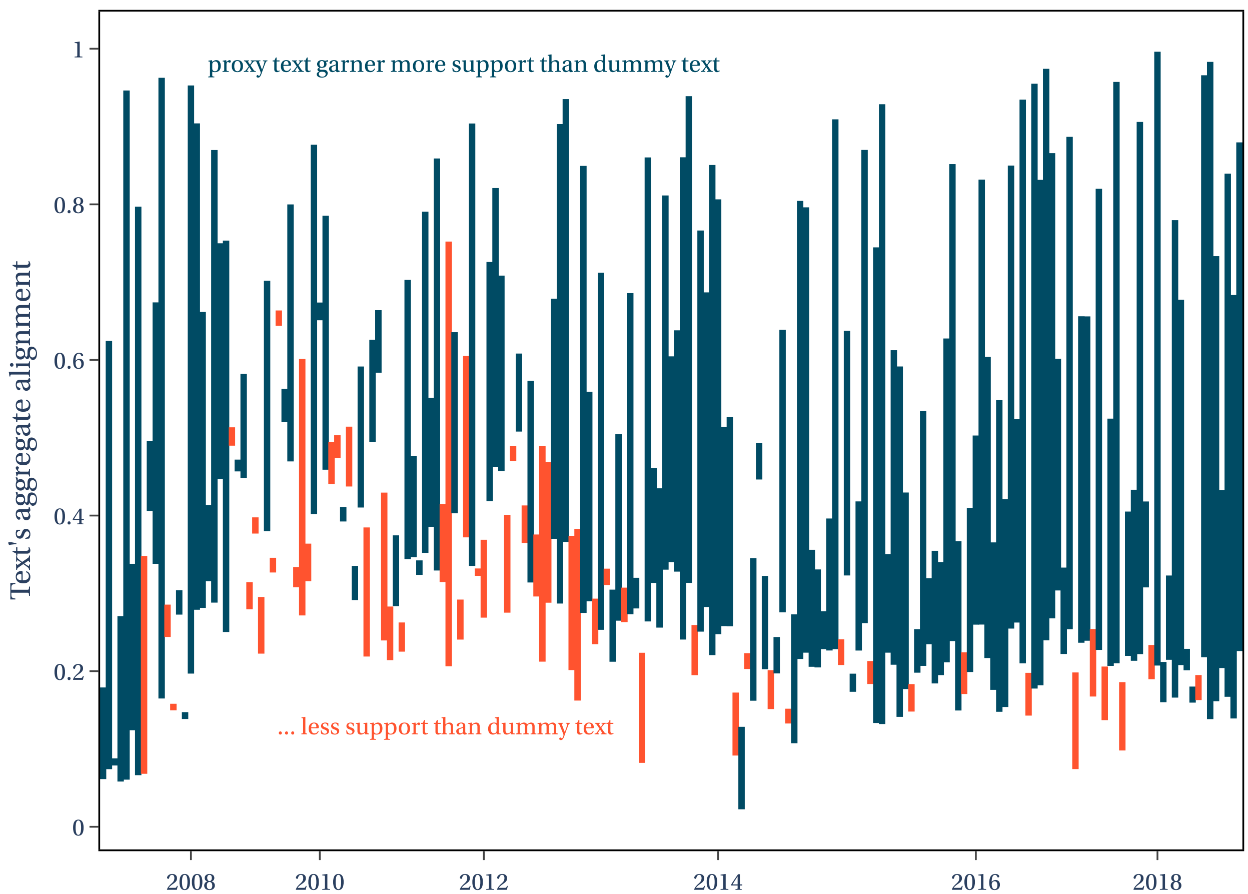
6.2 Robustness to a dummy attack text

How much is the attack text's alignment different from something gibberish? Would we see lower alignment if the activists say something illegible? To test the counterfactual, I stitch an attack text from all the attack text in my sample. I use stitched text as the counterfactual dummy text instead of picking random words from the sample for two reasons: (i) randomness loses the replicability aspect (ii) a random bag of words does not make legible higher-order phrases, which are part of the up to five-word phrase SVR model. To create the stitched text, I sort the 533 attack texts based on the date. I standardize each text, similar to the process described in Section 3.1. For each attack, I calculate the length of words to take, $numw_a$, as the total number of words in the attack text divided by 533. Next, I take the first $numw_1$ words from the first attack text for the stitched text. From the second attack onward, I start at $numw_t \times (t-1)$ and add the next $numw_t$ words to the stitched text. I use the stitched text of 997 words as the dummy text for each attack.

Although attacks pandering to larger shareholders are pervasive in activism, their empirical relevance is unclear. Figure 9 shows that such attacks are, in fact, widespread. The average aggregate alignment for the attack if it uses the attack text is 53%, compared to 28% for the dummy text. Of the 195 attacks, 145 attacks (or 74%), accomplish a higher alignment using the attack text. The difference between the alignments is significant at 1%. The difference in aggregate alignment increases over the years and could explain the higher success rates of activists.

On average, the attack text's aggregate alignment is higher by 25 percentage points, compared to a dummy text, indicating that activists fight on issues that matter to larger shareholders. Therefore, on average, activists get 1.7 percentage points (3/44*25) higher mutual fund support by carefully aligning the issues they discuss in attack texts, calculated based on Section 5.2. Similarly, the higher aggregate alignment increases the activist's likelihood of win-

Figure 9:
Attack text, compared to a dummy text, is better aligned with larger fund families.
The graph plots candlestick figures for the difference between aggregate alignment with mutual funds for attack text and the dummy text. At the attack level, the aggregate alignment is the sum of alignment with a fund family weighted by the fund family holdings. Alignment for a fund family is based on its voting choices in shareholder proposals. The sample is restricted to attacks where mutual funds with voting information own at least the average, 14.3%, of the targets' market cap over the 2004–2019 period. The dummy text is stitched from text parts of all the attack texts during the period described in Section 6.2. The endpoints of each stick mark the aggregate alignment for the actual attack text and the dummy text. The shade of the sticks indicates if the actual attack text, compared to the dummy text, has a higher aggregate alignment. I use SEC filings to get the attacks and CapitalIQ to identify the outcomes.



ning the attack by 8.4 percentage points (9.4/28*25), calculated based on Section 5.3.

6.3 Robustness to a non-machine learning method

Even though the SVR method is interpretable and based on the fund family's voting choices, it is complex to track. In my analysis, the SVR has coefficients for more than nine thousand phrases, which decide the attack text's alignment. As such, it is cumbersome to keep track of all the moving parts. In this section, I check whether my result of positive association between an attack text's alignment and fund family ownership holds if I use a simpler non-machine learning method to measure of alignment.

While the ISS voting database does not provide texts of proposals, it gives a one-line description of the proposal, usually the heading. I categorize the most common descriptions into 25 proposal types, including director election, governance, sustainability, etc. Starting from the most common proposal types, I am able to classify 90% of the shareholder proposals into one of the 25 proposal types. Appendix F.1 lists the classification of proposal descriptions into different types. I use the method described in Section 5.2 to get proposals related to an attack. Usually, an attack has more than one associated shareholder proposal: my sample consists of 78 attacks that went to a voting stage, involving, on average, six proposals.

I define a fund family's alignment with an attack proposal as the fraction of relevant shareholder proposals in which the fund family voted against management recommendation. To gather the relevant shareholder proposals for an attack proposal, I look for shareholder proposals of the same type as the attack proposal and have meeting dates within two years prior to the attack's beginning date. To get the attack's alignment with the fund family preferences, I average alignment of attack proposals within an attack. Table 11 shows the result of the re-evaluation of Section 4.1 using proposal-type based measurement of alignment. A one-standard-deviation increase in fund family holdings of targeted shares, which is approximately 0.66 percentage point relative to an average of 0.1%, is associated with around a 0.6 percentage point increase in the attack text's alignment with the fund family preferences. The coefficients

Table 11:
Activists include proposal types that are well-aligned with larger shareholders
This table reports estimates of regression of an attack text's alignment with fund family preferences on the fund families' holdings in targets. Specifically, I estimate:

$$\widehat{Align}_{a,f} = \beta Holding_{a,f} + \delta_a + \delta_f + \epsilon_{f,a}$$

where $\widehat{Align}_{a,f}$ is the predicted alignment of attack text, *a*, with the fund family, *f*, preferences. *Align* is calculated based on the family's voting in shareholder proposals that of the same type as attack proposals. *Holding* is the percent of equity the fund family owns of the target before the attack, obtained from the CRSP database. δ_a and δ_f represent attack level and fund family level fixed effects, respectively. The sample consists of attacks, identified using attack filings, that went to a voting stage over the 2004–2019 period. Corresponding fund families include all the families that have voted in at least a hundred shareholder proposals in the two years prior to the attack. The independent variable is scaled by the standard deviation of the underlying variable, meaning the coefficient can be interpreted as the effect of a one-standard-deviation change in the determinant. Standard errors, $\epsilon_{a,f}$, are clustered at the attack level, and *t*-statistics are reported in brackets below the coefficient estimates. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

	Attack text's alignment with fund family preferences (simpler method)			
	(1)	(2)	(3)	(4)
Fraction of target mcap	0.0270***	0.0288**	0.0038	0.0064*
held by fund family	[10.67]	[9.04]	[1.16]	[1.97]
Attack FE		Yes		Yes
Fund family FE			Yes	Yes
Observation	13,328	13,328	13,326	13,326
R^2	0.008	0.136	0.370	0.494

are comparable to numbers in Table 5. The results indicate that activists selectively use proposal types on which the larger shareholders have voted against the management.

6.4 Robustness to changing parameters

Results in the paper are also robust to various specification choices. In Section 5.3, I use a cutoff for mutual fund ownership to define the ownership dummy. The dummy is one if mutual funds with voting information own more than the average ownership of 14.3%. Appendix Figure 12 shows that the results in Section 5.3 hold for different cutoff parameters. For the

SVR method, I make subjective choices in terms of parameters used: (i) n-gram length = 5, (ii) threshold for excluding words with higher frequency = 0.7, (iii) minimum number of voting observations = 100, and (iv) window of shareholder proposals = 2 years. Appendix Figure 13 shows coefficients with a 95% confidence interval for Equation 4. The coefficients are significant for changing parameters on either side of the respective cutoffs. Thus, the text-based voting prediction is rooted in fund families' proxy guidelines and is insensitive to changing parameters.

7 Conclusion

In this paper, I examine the role of shareholder preference in shaping activists' campaigns, tactics, and successes. In their proxy communications, I find that activists use language to garner support from fund families that own a larger share in the targeted firm. I use fund family voting history on shareholder proposals to measure a fund family's preference and, subsequently, the attack text's alignment with the fund family. The alignment is positively associated with funds' holdings in the target. For every percent increase in holdings, the attack text's alignment with fund family preferences increases by 0.7 percentage points. The results suggest that activists and targets selectively raise issues that are important to larger shareholders.

The use of tailored campaigns partly explains activists' increased success and amicable relations with long-term shareholders in recent years. I find an increased use of well-aligned attack text over the years. There is evidence of a learning curve for the activists, where they tailor their campaigns better as they interact more with fund families.

The finding echoes the concern that a few shareholders wield disproportionate power over the direction of corporate America. WSJ (2020a) notes that "markets are shifting from harnessing the wisdom of crowds to the wisdom of a handful of influential money-management executives." As a prevention mechanism against anti-competitive influence, the SEC has different filing requirements based on the investors' desire to influence. While an institution

looking for change has to file the stringent beneficial ownership form 13D and disgorge profit on trades for six months (Section 16B), an institution in the ordinary course of business needs to file a much less stringent form 13G (Morley 2018). However, my findings suggest that the activists provide a channel via which fund preferences influence corporate governance.

Institutions that manage mutual funds are an essential component of financial markets. However, they are restricted, both legally and incentive-wise, from engaging with managers in an openly confrontational way. More often than not, they are accused of being excessively deferential toward managers of portfolio companies, who affect the private interests of the index fund manager(Bebchuk and Hirst 2019). Lund (2017) even argues that index mutual funds should abstain from voting, leaving decisions to those with an incentive to be informed. In this paper, I find that the hedge funds pander to fund families by focusing on issues that are important to the families. Attack texts that are well-aligned with a fund family's preferences catch the fund family's attention and votes. A one-standard-deviation increase in attack text alignment, which is around 40 percentage points, is associated with a 23% increase in fund family attention and a 6% increase in actual activist support. These attacks are also more likely to end up in favor of the activists. For a one-standard-deviation increase in aggregate attack text alignment, which is approximately 28 percentage points, the likelihood of activists winning the attack increases by 15%, a significant improvement.

Although I validate my findings based on proxy guidelines texts, a non-machine learning based approach, and differing parameters, I lack counterfactual data. What if the activists used a different attack text? How would that have turned out? Or would an exogenous shock to target ownership during an attack change what activists discuss? Unfortunately, the data set for activism does not provide the depth to do that. I also look at just one of the many dimensions in which persuasion manifests in activism. Examples of other persuasion dimensions, which this paper does not explore, include behind the scenes meetings between shareholders and targets' management, activists' media interactions, etc. Future research that assesses exogenous shocks to the interactions between parties and other modes of communication could

shed more light on ways persuasion plays out in shareholder activism.

References

Aggarwal, R., P. a. C. Saffi, and J. Sturgess (2015). The Role of Institutional Investors in Voting: Evidence from the Securities Lending Market. *The Journal of Finance 70*(5), 2309–2346.

Alchian, A. A. (1950, June). Uncertainty, Evolution, and Economic Theory. *Journal of Political Economy 58*(3), 211–221.

Alchian, A. A. and H. Demsetz (1972). Production, Information Costs, and Economic Organization. *The American Economic Review 62*(5), 777–795.

Alexander, C. R., M. A. Chen, D. J. Seppi, and C. S. Spatt (2010, November). Interim news and the role of proxy voting advice. *Review of Financial Studies 23*(12), 4419–4454.

Allaire, Y. (2013, January). The Troubling Case of Proxy Advisors: Some Policy Recommendations. SSRN Scholarly Paper ID 2282617, Social Science Research Network, Rochester, NY.

Appel, I. R., T. A. Gormley, and D. B. Keim (2019, July). Standing on the Shoulders of Giants: The Effect of Passive Investors on Activism. *The Review of Financial Studies 32*(7), 2720–2774.

Ashraf, R., N. Jayaraman, and H. E. Ryan (2012). Do Pension-Related Business Ties Influence Mutual Fund Proxy Voting? Evidence from Shareholder Proposals on Executive Compensation. *The Journal of Financial and Quantitative Analysis 47*(3), 567–588.

Bainbridge, S. M. (2005). Director Primacy and Shareholder Disempowerment Response to Increasing Shareholder Power. *Harvard Law Review 119*(6), 1735–1758.

Bauguess, S. W., J. Cooney, and K. W. Hanley (2013). Investor demand for information in newly issued securities. *SSRN Electronic Journal*.

Bebchuk, L. A., A. Cohen, and S. Hirst (2017, August). The Agency Problems of Institutional Investors. *Journal of Economic Perspectives 31*(3), 89–102.

Bebchuk, L. A. and S. Hirst (2019, December). Index Funds and the Future of Corporate Governance: Theory, Evidence, and Policy. Working Paper 26543, National Bureau of Economic Research.

Bessi, A., M. Coletto, G. A. Davidescu, A. Scala, G. Caldarelli, and W. Quattrociocchi (2015, February). Science vs Conspiracy: Collective Narratives in the Age of Misinformation. *PLOS ONE 10*(2), e0118093.

BlackRock (2015). Larry fink's letter to CEOs, 2015. http://www.shareholderforum.com/access/Library/20150331_BlackRock.pdf.

Bodnaruk, A. and M. Rossi (2016, April). Dual ownership, returns, and voting in mergers. *Journal of Financial Economics 120*(1), 58–80.

Bolton, P., T. Li, E. Ravina, and H. Rosenthal (2020, August). Investor ideology. *Journal of Financial Economics 137*(2), 320–352.

Boyson, N. M., N. Gantchev, and A. Shivdasani (2017, October). Activism mergers. *Journal of Financial Economics 126*(1), 54–73.

Boyson, N. M. and R. M. Mooradian (2011, July). Corporate governance and hedge fund activism. *Review of Derivatives Research 14*(2), 169–204.

Bradley, M., A. Brav, I. Goldstein, and W. Jiang (2010, January). Activist arbitrage: A study of open-ending attempts of closed-end funds. *Journal of Financial Economics 95*(1), 1–19.

Brav, A., A. Dasgupta, and R. D. Mathews (2019, May). Wolf Pack Activism. SSRN Scholarly Paper ID 2529230, Social Science Research Network, Rochester, NY.

Brav, A., W. Jiang, and H. Kim (2010). *Hedge Fund Activism: A Review*. Now Publishers Inc.

Brav, A., W. Jiang, and H. Kim (2015a, October). The Real Effects of Hedge Fund Activism: Productivity, Asset Allocation, and Labor Outcomes. *The Review of Financial Studies 28*(10), 2723–2769.

Brav, A., W. Jiang, and H. Kim (2015b). Recent Advances in Research on Hedge Fund Activism: Value Creation

and Identification. *Annual Review of Financial Economics 7*(1), 579–595.

Brav, A., W. Jiang, T. Li, and J. Pinnington (2018, March). Picking Friends Before Picking (Proxy) Fights: How Mutual Fund Voting Shapes Proxy Contests. SSRN Scholarly Paper ID 3101473, Social Science Research Network, Rochester, NY.

Brav, A., W. Jiang, S. Ma, and X. Tian (2018, November). How does hedge fund activism reshape corporate innovation? *Journal of Financial Economics 130*(2), 237–264.

Brav, A., W. Jiang, F. Partnoy, and R. Thomas (2008). Hedge Fund Activism, Corporate Governance, and Firm Performance. *The Journal of Finance 63*(4), 1729–1775.

Bubb, R. and E. Catan (2018, February). The Party Structure of Mutual Funds. SSRN Scholarly Paper ID 3124039, Social Science Research Network, Rochester, NY.

Buchanan, B. G., J. M. Netter, A. B. Poulsen, and T. Yang (2012). Shareholder Proposal Rules and Practice: Evidence from a Comparison of the United States and United Kingdom. *American Business Law Journal 49*(4), 739–803.

Butler, A. W. and U. G. Gurun (2012, August). Educational Networks, Mutual Fund Voting Patterns, and CEO Compensation. *The Review of Financial Studies 25*(8), 2533–2562.

Cai, J., J. L. Garner, and R. A. Walkling (2009). Electing Directors. *The Journal of Finance 64*(5), 2389–2421.

Cai, J. and R. A. Walkling (2011). Shareholders' Say on Pay: Does It Create Value? *The Journal of Financial and Quantitative Analysis 46*(2), 299–339.

Chou, J., L. Ng, and Q. Wang (2011, December). Are better governed funds better monitors? *Journal of Corporate Finance 17*(5), 1254–1271.

Clifford, C. P. (2008, September). Value creation or destruction? Hedge funds as shareholder activists. *Journal of Corporate Finance 14*(4), 323–336.

Coffee Jr, J. C. and D. Palia (2016, February). The Wolf at the Door: The Impact of Hedge Fund Activism on Corporate Governance. *Annals of Corporate Governance 1*(1), 1–94.

Cvijanović, D., A. Dasgupta, and K. E. Zachariadis (2016). Ties That Bind: How Business Connections Affect Mutual Fund Activism. *The Journal of Finance 71*(6), 2933–2966.

Davis, G. F. and E. H. Kim (2007, August). Business ties and proxy voting by mutual funds. *Journal of Financial Economics 85*(2), 552–570.

Del Vicario, M., A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi (2017, January). Modeling confirmation bias and polarization. *Scientific Reports 7*(1), 40391.

Del Vicario, M., G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, and W. Quattrociocchi (2016, December). Echo Chambers: Emotional Contagion and Group Polarization on Facebook. *Scientific Reports 6*(1), 37825.

Drake, M. S., D. T. Roulstone, and J. R. Thornock (2015, March). The determinants and consequences of information acquisition via EDGAR. *Contemporary Accounting Research 32*(3), 1128–1161.

Drucker, H., Christopher, L. Kaufman, A. J. Smola, and V. Vapnik (1997). Support vector regression machines. *Nips.cc*, 155–161.

Edmans, A. and C. G. Holderness (2017, January). Chapter 8 - Blockholders: A Survey of Theory and Evidence. In B. E. Hermalin and M. S. Weisbach (Eds.), *The Handbook of the Economics of Corporate Governance*, Volume 1 of *The Handbook of the Economics of Corporate Governance*, pp. 541–636. North-Holland.

Ertimur, Y., F. Ferri, and D. Oesch (2013). Shareholder Votes and Proxy Advisors: Evidence from Say on Pay. *Journal of Accounting Research 51*(5), 951–996.

Fidelity (2019). Fidelity Securities Fund Prospectus. https://www.sec.gov/Archives/edgar/data/754510/000137949119001430/filing989.htm.

FlaggStreet (2007). Flagg Street Capital LLC Form PREC14A, number 0000921895-07-000903 | SEC Filings API. https://sec-api.com/filing/acc/0000921895-07-000903.

Fos, V. and M. Tsoutsoura (2014, November). Shareholder democracy in play: Career consequences of proxy contests. *Journal of Financial Economics 114*(2), 316–340.

FrontFour (2013). FrontFour Capital Group LLC Form DFAN14A, number 0000921895-13-000151 | SEC Filings API. https://sec-api.com/filing/acc/0000921895-13-000151.

FT (2016). Ackman says ick to etfs. https://ftalphaville.ft.com/2016/01/27/2151315/ackman-says-ick-to-etfs/.

Gantchev, N. (2012, August). The costs of shareholder activism: Evidence from a sequential decision model. *Journal of Financial Economics 107*.

Gillan, S. L. and L. T. Starks (2000, August). Corporate governance proposals and shareholder activism: The role of institutional investors. *Journal of Financial Economics 57*(2), 275–305.

Gormley, T. A. and M. Jha (2020, October). Bonds Lie in the Portfolio of the Beholder: Do Bonds Affect Equity Monitoring? SSRN Scholarly Paper ID 3719187, Social Science Research Network, Rochester, NY.

Greenwood, R. and M. Schor (2009). Investor activism and takeovers. *Journal of Financial Economics 92*(3), 362–375.

Gu, Z. and C. Zhang (2020). Living with the frenemy: Common ownership and hedge fund activism.

Harford, J., D. Jenter, and K. Li (2011, January). Institutional cross-holdings and their effect on acquisition decisions. *Journal of Financial Economics 99*(1), 27–39.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *Springer Series in Statistics the Elements of Statistical Learning Data Mining, Inference, and Prediction Second Edition*.

He, Y. and T. Li (2018, April). The Benefits of Friendship in Hedge Fund Activism. SSRN Scholarly Paper ID 2794709, Social Science Research Network, Rochester, NY.

He, Y. E., B. Kahraman, and M. Lowry (2018). Mutual fund voting on environmental and social proposals. *Available at SSRN*.

Hemphill, C. S. and M. Kahan (2019). The Strategies of Anticompetitive Common Ownership. *Yale Law Journal 129*, 1392.

Icahn (2013). Icahn Carl DFAN14A Filing. https://www.sec.gov/Archives/edgar/data/921669/0000928464-13-000275.txt.

Iliev, P., J. Kalodimos, and M. Lowry (2020, July). Investors' Attention to Corporate Governance. SSRN Scholarly Paper ID 3162407, Social Science Research Network, Rochester, NY.

Iliev, P. and M. Lowry (2015, February). Are Mutual Funds Active Voters? *The Review of Financial Studies 28*(2), 446–485.

InstitutionalInvestor (2006). The fairest of them all. https://www.institutionalinvestor.com/article/b150nqg3ffc5zk/the-fairest-of-them-all.

Jøsang, A., W. Quattrociocchi, and D. Karabeg (2011). Taste and Trust. In I. Wakeman, E. Gudes, C. D. Jensen, and J. Crampton (Eds.), *Trust Management V*, IFIP Advances in Information and Communication Technology, Berlin, Heidelberg, pp. 312–322. Springer.

Kamenica, E. and M. Gentzkow (2011, October). Bayesian Persuasion. *American Economic Review 101*(6), 2590–2615.

Kedia, S., L. T. Starks, and X. Wang (2020, March). Institutional Investors and Hedge Fund Activism. SSRN Scholarly Paper ID 3560537, Social Science Research Network, Rochester, NY.

Klein, A. and E. Zur (2009). Entrepreneurial Shareholder Activism: Hedge Funds and Other Private Investors. *The Journal of Finance 64*(1), 187–229.

Kogan, S., D. Levin, B. R. Routledge, J. S. Sagi, and N. A. Smith (2009, May). Predicting risk from financial reports with regression. In *The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL '09, Boulder, Colorado, pp. 272–280. Association for Computational Linguistics.

Kubat, M., S. Matwin, et al. (1997). Addressing the curse of imbalanced training sets: One-sided selection. In

Icml, Volume 97, pp. 179–186. Nashville, USA.

Li, Z. F., S. Patel, and S. Ramani (2019). The role of mutual funds in corporate social responsibility. *SSRN Electronic Journal*.

Loughran, T. and B. McDonald (2017). The use of EDGAR filings by investors. *Journal of Behavioral Finance 18*(2), 231–248.

Lund, D. S. (2017). The Case against Passive Shareholder Voting. *Journal of Corporation Law 43*, 493.

Malenko, N. and Y. Shen (2016, December). The Role of Proxy Advisory Firms: Evidence from a Regression-Discontinuity Design. *The Review of Financial Studies 29*(12), 3394–3427.

Manela, A. and A. Moreira (2017). News implied volatility and disaster concerns. *Journal of Financial Economics 123*(1), 137–162.

Matvos, G. and M. Ostrovsky (2008, September). Cross-ownership, returns, and voting in mergers. *Journal of Financial Economics 89*(3), 391–403.

Matvos, G. and M. Ostrovsky (2010, October). Heterogeneity and peer effects in mutual fund proxy voting. *Journal of Financial Economics 98*(1), 90–112.

McCahery, J. A., Z. Sautner, and L. T. Starks (2016, November). Behind the scenes: The corporate governance preferences of institutional investors. *The Journal of Finance 71*(6), 2905–2932.

McCloskey, D. and A. Klamer (1995). One Quarter of GDP is Persuasion. *The American Economic Review 85*(2), 191–195.

Mietzner, M. and D. Schweizer (2014, April). Hedge funds versus private equity funds as shareholder activists in Germany — differences in value creation. *Journal of Economics and Finance 38*(2), 181–208.

Morgan, A., A. Poulsen, J. Wolf, and T. Yang (2011, September). Mutual funds as monitors: Evidence from mutual fund voting. *Journal of Corporate Finance 17*(4), 914–928.

Morley, J. D. (2018). Too Big to Be Activist. *Southern California Law Review 92*, 1407.

Norli, O., C. Ostergaard, and I. Schindele (2014, October). Liquidity and shareholder activism. *The Review of Financial Studies 28*(2), 486–520.

NPR (2017). Why bill ackman sees activist investing as a moral crusade. https://www.nprillinois.org/post/why-bill-ackman-sees-activist-investing-moral-crusade.

NYTimes (2019, August). Opinion | Why Isn't Your Mutual Fund Sticking Up for You? https://www.nytimes.com/2019/08/23/opinion/mutual-funds-shareholder-activism.html.

Ramius (2009). Ramius Schedule 14A Form. https://www.sec.gov/Archives/edgar/data/25354/000092189509001485/prec14a06297038_05222009.htm.

RelBanks (2017). Top Asset Management Firms. https://www.relbanks.com/rankings/largest-asset-managers.

Reuters (2018). Blackrock's fink learns to live with activist investors. https://www.reuters.com/article/us-investment-summit-fink-shareholders/blackrocks-fink-learns-to-live-with-activist-investors-idUSKBN1DD2B6.

Rothberg, B. and S. Lilien (2006, January). Mutual Funds and Proxy Voting: New Evidence on Corporate Governance. *Journal of Business & Technology Law 1*(1), 157.

Schwartz-Ziv, M. and E. Volkova (2020, June). Is Blockholder Diversity Detrimental? SSRN Scholarly Paper ID 3621939, Social Science Research Network, Rochester, NY.

SEC (2019). Commission Guidance Regarding Proxy Voting Responsibilities of Investment Advisers. https://www.sec.gov/rules/interp/2019/ia-5325.pdf.

Shleifer, A. and R. W. Vishny (1986, June). Large Shareholders and Corporate Control. *Journal of Political Economy 94*(3, Part 1), 461–488.

StateStreet (2020). Summary of material changes to state street global advisors' 2020 proxy voting and en-

gagement guidelines. https://www.ssga.com/library-content/pdfs/global/proxy-voting-and-engagement-guidelines.pdf.

Swire, B., U. K. H. Ecker, and S. Lewandowsky (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition 43*(12), 1948–1961.

TheStreet (2017). How Index Funds Are Taking Over And What It Means For Activist Investing. https://www.thestreet.com/markets/mergers-and-acquisitions/how-index-funds-could-turbocharge-activism-14229514.

Vanguard (2010). Vanguard Index Fund Form 485BPOS. https://www.sec.gov/Archives/edgar/data/36405/0000932471-10-001958.txt.

Vanguard (2018). Vanguard Charlotte Funds Prospectus. https://www.sec.gov/Archives/edgar/data/1532203/000093247118006540/globalcreditbondfund485a.htm.

Vanguard (2019). Vanguard Index Fund Form 485BPOS. https://www.sec.gov/Archives/edgar/data/36405/0000932471-19-006954.txt.

Wong, Y. T. F. (2019, November). Wolves at the Door: A Closer Look at Hedge Fund Activism. *Management Science 66*(6), 2347–2371.

WSJ (2018). BlackRock, Calpers Want Exchanges to Clamp Down on Dual-Class Shares. https://www.wsj.com/articles/blackrock-and-calpers-to-stock-exchanges-clamp-down-on-dual-class-shares-1540394503.

WSJ (2019). Vanguard Is Handing Over Some of Its Voting Power. https://www.wsj.com/articles/vanguard-is-handing-over-some-of-its-voting-power-11556190120.

WSJ (2020a). Replacing the wisdom of crowds with the wisdom of fink. https://www.wsj.com/articles/replacing-the-wisdom-of-crowds-with-the-wisdom-of-fink-11579429800.

WSJ (2020b). WSJ News Exclusive | How Investing Giants Gave Away Voting Power Ahead of a Shareholder Fight. https://www.wsj.com/articles/how-investing-giants-gave-away-voting-power-ahead-of-a-shareholder-fight-11591793863.

Xie, J. and J. Gerakos (2020, May). The Anticompetitive Effects of Common Ownership: The Case of Paragraph IV Generic Entry. *AEA Papers and Proceedings 110*, 569–572.

Appendices

A Data collection

A.1 Assigning proposal text to ISS voting data

Voting records of the fund family, at the mutual fund level, are available from the ISS. I aggregate mutual fund voting information into fund family voting data, based on the names of the mutual funds, mergers and acquisition, and investment relationships among mutual fund institutions. During the period 2003–2018, I have 359 fund families who have voted in at least a hundred proposals. These fund families voted on a total of 10,679 unique shareholder proposals. For the text of the proposals, I use DEF14A filings, which are available at the EDGAR system via the SEC. The system provides indexes to all public filings, including CIK, type of form, filing date, and weblink.

I match proposals in the ISS voting database to the text available in the DEF14A filings. The voting data provides a record date, meeting date, proposal item number, and a short description of the proposal. To make a suitable match, I start by slicing the shareholder proposal for a particular CIK and subsequently for a specific meeting date. To get a list of potential text matches for a proposal in voting data, I employ a two-step process. First, I gather all the proposals for a particular CIK on a meeting date in the voting data. Next, I slice the SEC index file for the particular CIK, DEF14A filing, and filing date between the record and meeting date. The average number of proposals for a firm on a meeting date is 2.17. Usually, proposals about director elections are grouped as one in DEF14A filings. Therefore, for searching in DEF14A, I combine all the director election proposals into a single proposal.

I parse the DEF14A HTML file using Beautiful Soup python package. I remove all the tables, white space, accented characters, and non-UTF encoding. I also filter out the first 75, which has filer information, and the last 75 lines, which are often errors from PDF to HTML conversion, from the filings. Once I have the clean DEF14A text, I look for sections of the filing

that correspond to the specific proposal. To get the starting line for a proposal, I assign a score to each line of the DEF14A based on how likely it matches the ISS proposal description and item number. I choose the line with the maximum score. I assign higher scores if the line (i) is uppercase, (ii) contains words such as proposal, number, no., item, etc. (iii) contains the same words as it appears in ISS description (iv) has less than 80 characters (v) contains the same number as ISS item number. Sometimes the proposals are written in two lines - the first line containing item number and the second containing the description. To take this into account, I repeat the same process by combining two consecutive lines and checking the score improvement.

To find the starting line for the next proposal, I begin five lines after the previous proposal's start line. Proposals in DEF14A are typically sequentially put; thus, I choose the ending for a proposal as two lines before the starting of the next one. To get the last proposal's ending line, I begin five lines after the start and look for the phrase "The Board of Directors recommends." In there are no matches, I take the ending line as fifty lines after the starting line. I assign the text between the starting line and the ending line in DEF14A to each proposal. For director election proposals, which generally have one proposal for all the nominated directors, I choose paragraphs between the starting line and the ending line that contains the name of the director listed in the voting database.

Out of the 10,679 shareholder proposals, I assign text to 6,176 proposals. The difference in numbers is because (i) the ISS data includes shareholder proposal for companies across the globe, while SEC filings are done by US-based companies (ii) some of the proposals are written in a nonstandard format, which makes parsing them precisely difficult (iii) given the goal of this study is to analyze the text, I limit the sample to proposals, where I am able to match text information with reasonable confidence, and which has more than 30 words.

A.2 Extracting text associated with an attack

To get the attack text, I look for form DEFC14A, DFAN14A, and PREC14A filed by the investment firm and parse filer and subject company CIKs. Two fields characterize the proxy filings associated with attacks: (i)*FILED BY*, containing the activist information, and (ii)*SUBJECT COMPANY*, containing information on the targeted firm (target). To get information on these attack text, I begin with the filer particulars. Every institutional investment managers with at least $100 million in equity assets under management are required to file a 13F form with SEC. Thus, I include only those filers that have filed 13F-HR, 13F-NT, or 13F-E form to make a list of all the investment firms. Activists must file with the SEC if they discuss material information even if the information is not part of a campaign. I filter out filings that (i) do not contain text, (ii) refer to an external exhibit document, and (iii) are related to merger and acquisition, litigation, or banter (Icahn 2013). I remove duplicate filings, which are usually the same document filed by the subject company for easier access to shareholders.

I get a total of 4,159 proxy filings related to attacks, which includes 290 DEFC14A, 3,484 DFAN14A, and 385 PREC14A filings. I combine these proxy filings if they are less than 180 days apart and have the same activist and target. However, in two cases, I combine filings that are more than 180 days apart - the 2006 attack on Sunset Financial and the 2018 attack on Alpine dividend fund. I get 533 confrontational proxy attacks, over the 2004–2019 period, with an average of eight filings per attack. The proxy filings contain information related to activist identification, activist's message to shareholders, voting procedure, activist's holdings in the target firm, other legal disclosure. Sometimes the activist also discusses their portfolio, past activism success, etc. I parse out the activist's message to shareholders from each filing and combine the messages across filing to get the attack text. To parse out the message part, I look for cues that begin and end a message. Table 12 lists the ten most common cues.

Table 12:
Top ten cues to parse the message in attack filings
This table reports a subset of cues to get the message to shareholders, from an attack filing. [ACTIVIST] ([TARGET]) is a placeholder for the name of activist (target), gathered from identification section in the attack filing.

Message Begin Cue	Message End Cue
Reasons for the solicitation	Sincerely yours
Ladies and Gentlemen	Warm regards
Dear Fellow Shareholder	Sincerely
Dear Board of Directors	Best
[ACTIVIST] is seeking your support for	Please sign date and return the gold proxy card today
The following is the text of a press release issued by [ACTIVIST]	Security holders are advised to read the proxy statement and other documents related to the solicitation of proxies
confirms intention to nominate [ACTIVIST]	Urge you to vote your shares on the green proxy card
find proxy materials for the important annual meeting of [TARGET]	Please address any correspondence to [ACTIVIST]
being furnished to you the stockholders of [TARGET]	For further information including full biographies of our management team
soliciting proxies from holders of shares of [TARGET]	Any other relevant documents are available at no charge on the secs website

A.3 Processing fund's information acquisition via EDGAR

The search traffic data for SEC.gov covers the period from February 2003 through June 2017. EDGAR log file data set includes information on visitor's IP address, date, timestamp, CIK, and filing document's accession number. The IP addresses in the dataset are in version 4 (IPv4) format, which defines an IP address as a 32-bit number separated in four 8-bit numbers. A dot separates each 8-bit number, and the number between the dots could be between 0 and 255 ($2^8 - 1$). So a specific IP address, let's say BlackRock's, looks like 199.253.64.128. However, the last octet of the IP address in log files is replaced with alphabets, in a way to preserve the uniqueness of the IP address without revealing the full identity of the visitor. Thus, if Blackrock accesses the SEC.gov website from the IP address, the log file will show an entry 199.253.64.mns. In essence, the EDGAR log file dataset has a 24-bit (IP3) address for each EDGAR server activity. Fortunately, most of the fund family register large blocks of IP addresses. For example, BlackRock owns the IP addresses ranging from 199.242.6.0 to

199.242.6.255. As such, the IP3 addresses are often sufficient to pinpoint the registered fund family.

Loughran and McDonald (2017) advise separating EDGAR requests generated by robots from server requests by regular investors. I classify an IP address as a robot if it requests more than a thousand filings in a day. I remove IP addresses classified as robots for that particular day. To include only valid EDGAR activities, I follow Drake, Roulstone, and Thornock (2015) and exclude activities not related to governance research. I remove index pages (index.htm),icons (.ico), XML filings (.xml), and filings that are under 500 bytes in size. I also combine views by an IP address if they are less than five minutes apart and for the same filing.

The second part of my dataset is a lookup table from Digital Element, a geolocation data and services firm. The table contains the timestamp of IP addresses (IPv4) and registered organization name in December 2016. I use regular expressions, such as (.*blackrock.*) for BlackRock Financial Management, to get IPv4 associated with fund families. To assign IP3 blocks to fund families, I use a procedure similar to Iliev, Kalodimos, and Lowry (2020). If a fund family owns all or a subset of the IP3 address, and no other fund family owns an address from the IP3 block, I assign it to the fund family. If two or more fund families own a subset of IP3 block, I assign it to the family that contains the most IP address for the IP3 block. If two fund families own an equal number of IP addresses in an IP3 block, I drop those IP3 blocks. The chances of overestimating views from assigning an entire IP3 block to a fund family if they own a fraction of addresses is low, as it is unlikely for non-financial firms to access filings from SEC.gov.

Next, I look for the validity of IP3 blocks assigned to the fund family. The IP address to the organization name lookup table is a snapshot from December 2016. However, fund families sometimes change their underlying technology infrastructure and, in that process, register for different IP3 blocks. To make sure that I have credible IP3 blocks, I go back quarterly from December 2016 and see what fraction of holdings do fund family access through the EDGAR server. I use CRSP mutual fund data to get fund family holdings. If a fund family does not

access more than 1% of its holdings in two consecutive quarters, I stop including the fund family before the quarter. For example, Cambiar Investors accessed 1.9%, 3.3%, 0.0%, and 0.1% of its holdings in 2015Q4, 2015Q3, 2015Q2, and 2015Q1 respectively. Therefore, I exclude Cambiar Investors from my sample before June 2015. Subsequently, I match valid IP3 blocks from the organization lookup table with IP3 from EDGAR log files.

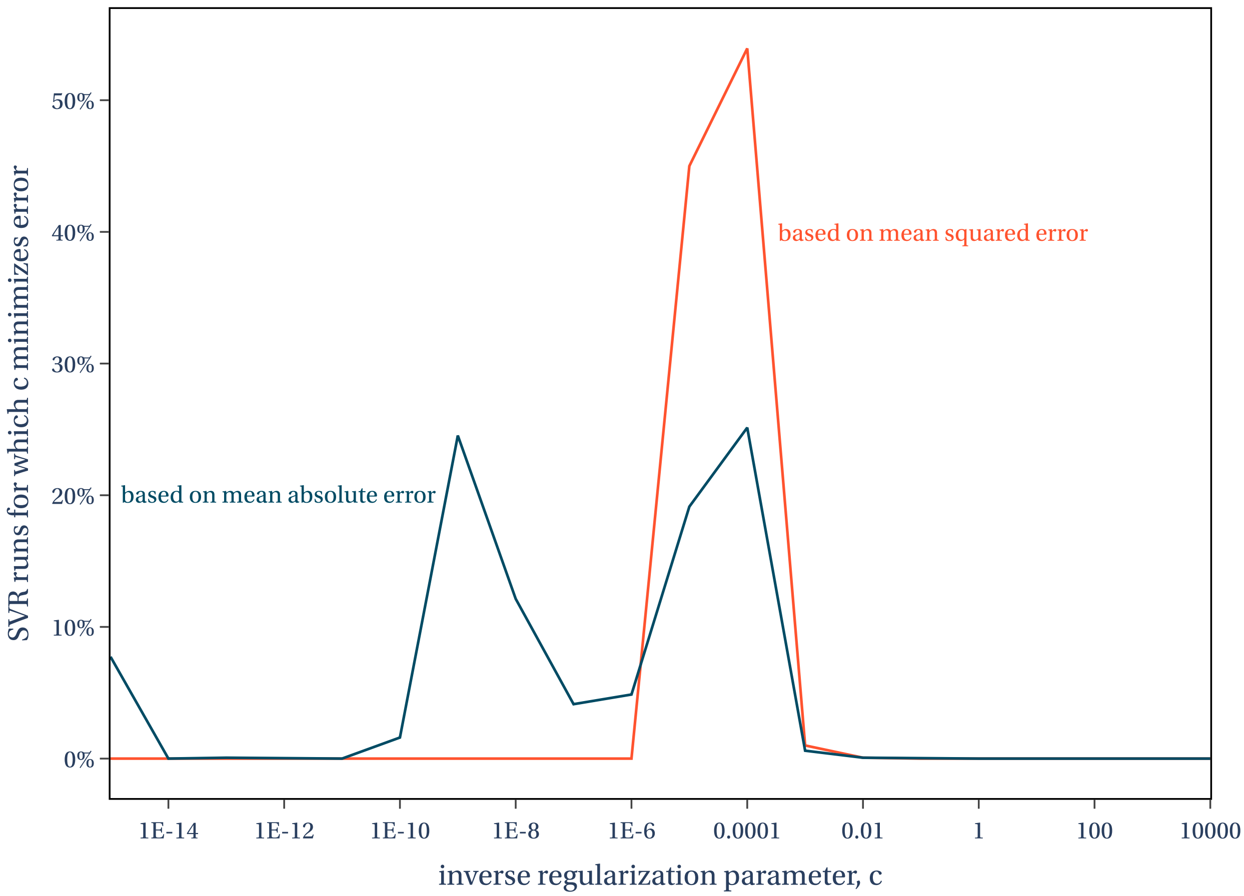
I identify attack documents based on the accession number of the filing in log files and SEC's index files. To measure the number of times a fund family accessed attack related filings, I aggregate views for attack's documents during the attack period, defined as the period from the first attack filing to 30-day after the last attack filing. The fund families' views, as measured from EDGAR log files, likely under-represent actual views. As mentioned in Bauguess, Cooney, and Hanley (2013), the EDGAR log files do not contain any requests for SEC filings from EDGAR's FTP site. Moreover, internet service providers cache frequently requested documents for future ease of reference. As such, requests for the same content that have been cached are not captured in the log file.

B Estimating SVR's parameter

SVR estimation requires the user to choose two hyperparameters, which control the trade-off between in-sample and out-of-sample fit: the ϵ insensitive zone and the inverse regularization parameter, c. I use an ϵ insensitive zone value of 0.001, i.e., the SVR method does not penalize the cost function if the difference between actual and predicted *Align* is less than 0.1% percent. I do not go more granular to improve computational efficiency, as differences in against management voting that are less than 0.1% does not mean much economically.

For the inverse regularization parameter, I run a horse-race amongst various values to obtain the lowest mean absolute error and mean squared error. I use a three-fold grid search algorithm for the proposal voting data to pick inverse regularization parameter, c, from 10^j, where j ranges from -15 to +4. I focus more on c below one as the strength of the regular-

Figure 10:
The inverse regularization parameter at 0.0001 minimizes out-of-sample errors.
The figure plots the percent of SVR runs for which an inverse regularization parameter reduces the out-of-sample mean absolute and mean squared error. The plot is based on SVR results of 25 randomly selected fund families at the end of each quarter over the 2004–2019 period (total run = 1500).



ization is inversely proportional to c. Figure 10 shows the best performing regularization parameter for the mean squared and the mean absolute error. The regularization parameter, c = 0.0001 has the lowest mean absolute error and mean squared error for 25% and 54% of my run sample. The run sample includes 25 randomly selected fund families each quarter over the 2004–2019 period, totalling to 1500 SVR runs. I use a three-fold cross-validation via GridSearch package to find the inverse regularization parameter with the lowest error.

C Examples of activists selectively using phrases

Figure 11 reports three examples of how activists focus on issues that matter to fund families with significant voting power. I choose these examples, as they have a variation in holdings between the three big fund families. In 2009, when Ramius LLC attacked CPI Corp, the focus of the attack was board members not having relevant industry experience. In the attack text, Ramius notes "experience board" 24 times (Ramius 2009). Incidentally, the experience of board members is important to Vanguard as well, which holds 2.8% of CPI shares.

Similarly, the 2007 Flagg Street Capital attack on Pomeroy Solutions was centered on an issue important to Fidelity, which owned 11.4% of Pomeroy shares. During the two years before the attack, Fidelity voted against management in 30% of shareholder proposals, compared to 25% by BlackRock and Vanguard, containing the phrase "personal benefits." Thus, an attack that discusses management's embezzlement would be closer to the preferences of Fidelity. The attack discussed how the Pomeroy family had run the company for personal benefit, including the transfer of the CEO position from David Pomeroy to his son (FlaggStreet 2007). Lastly, In 2013, FrontFour Capital discussed Ferro Corporation's deteriorating operating performance and how that has reflected on stock price (FrontFour 2013). Stock price performance was an important issue for both BlackRock and Vanguard, which together owned 10.3% of Ferro Corporation shares.

D Estimates for sub-sample of fund families that own target shares

E Limitations of text-based measure and mitigation

Certain limitations come with a text-based model. Many factors, including the content of an attack text, firm-specific performance, general economy, relationships between fund family

Figure 11:
Activists use phrases that will increase attack text's alignment with larger shareholders' preferences.
The bar chart shows the marginal increase in attack text's alignment with fund family preferences, if the activist uses one more instance of the phrase. The measure is derived from shareholder proposals voting in the two years before each attack. The vertical thin lines indicate the percent of target shares held by the fund family at the start of the attack. The x-axis mentions the key phrase, followed by Year Activist, Target tuple.

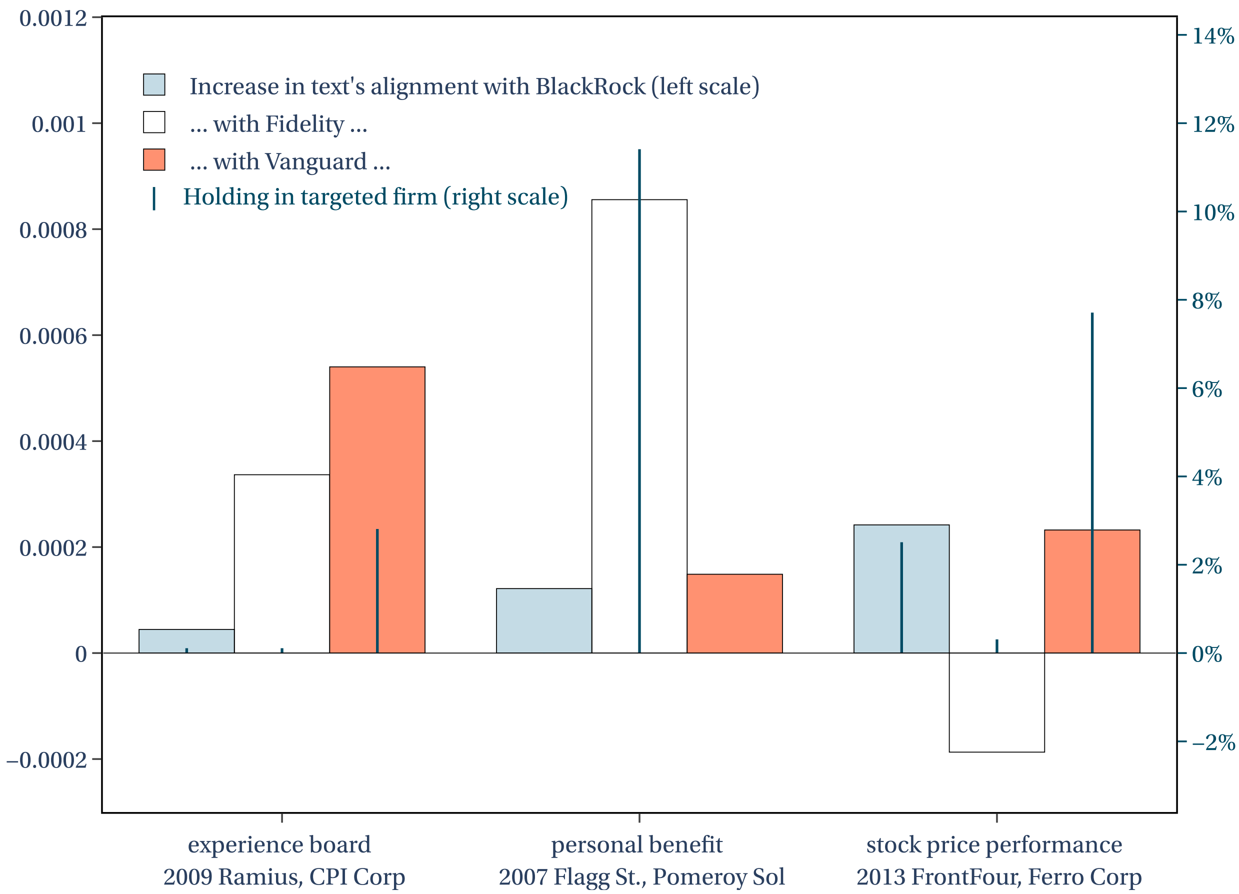


Table 13:
Activists tailor their communications to the preferences of large shareholders
This table reports estimates of regression of attack text's alignment with the fund family preferences on fund families' holdings in targets. Specifically, I estimate:

$$\widehat{Align}_{a,f} = \beta Holding_{a,f} + \delta_a + \delta_f + \epsilon_{f,a}$$

where $\widehat{Align}_{a,f}$ is the predicted alignment of attack text, *a*, with the fund family, *f*, preferences. *Holding* is the percent of equity the fund family owns of the target before the attack, obtained from the CRSP database. δ_a, and δ_f represent attack level and fund family level fixed effects, respectively. The sample consists of attacks, identified using SEC filings, over the 2004–2019 period. Corresponding fund families include all the invested families that have voted in at least a hundred shareholder proposals in the two years prior to the attack. The independent variable is scaled by the standard deviation of the underlying variable, meaning the coefficient can be interpreted as the effect of a one-standard-deviation change in the determinant. Standard errors, $\epsilon_{a,f}$, are clustered at the attack level, and t-statistics are reported in brackets below the coefficient estimates. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

	Attack text's alignment with fund family preferences			
	(1)	(2)	(3)	(4)
Fraction of target mcap	0.0144***	0.0112***	0.0125**	0.0065
held by fund family	[4.05]	[2.91]	[2.52]	[1.64]
Attack FE		Yes		Yes
Fund family FE			Yes	Yes
Observation	12,579	12,549	12,560	12,531
R^2	0.00	0.20	0.07	0.26

Table 14:
Fund families conduct more research on attacks tailored to their preferences
This table reports estimates of regression of fund family access of attack text filings on the attack text's alignment with fund family preferences. Specifically, I estimate:

$$View_{a,f} = \beta \widehat{Align}_{a,f} + \delta_a + \delta_f + \epsilon_{a,f}$$

where $View_{a,f}$ is the number of times a fund family, *f*, accessed attack filings, *a*, between the date the attack begins to 30 days after the attack ends. The attack's beginning (end) date is based on the first (last) date of attack filing by the activist. $\widehat{Align}_{a,f}$ is the attack text's alignment with fund family preferences. δ_a, and δ_f are attack level and are fund family level fixed effects. The sample consists of attacks, identified using attack filings, and corresponding fund families over the 2004–2019 period. Data for fund families' access of filings on SEC.gov is available via DERA. Columns (4), (5), and (6) control for fund family holdings in the target. Independent variables are scaled by the standard deviation of the underlying variable, meaning coefficients can be interpreted as the effects of a one-standard-deviation change in the determinant. Standard errors, $\epsilon_{a,f}$, are clustered at the attack level, and t-statistics are reported in brackets below the coefficient estimates. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% respectively.

	Number of times fund family viewed attack filings on SEC.gov					
	(1)	(2)	(3)	(4)	(5)	(6)
Attack text's	0.266***	0.298***	0.306***	0.246***	0.274***	0.302***
alignment	[4.52]	[3.77]	[4.14]	[4.21]	[3.54]	[4.08]
Fraction of target mcap				0.511***	0.584***	0.328***
held by fund family				[8.73]	[6.43]	[3.7]
Proxy fight FE		Yes	Yes		Yes	Yes
Fund family FE			Yes			Yes
Observation	10,232	10,207	10,198	10,232	10,207	10,198
R^2	0.002	0.147	0.226	0.009	0.156	0.228

and target, the reputation of activists, etc., play a role in how a fund family votes. Failure to control for such factors could introduce an omitted variable bias that confounds inferences. Moreover, the text-based measure is likely to predict a fraction of the whole voting variation, as shown in Section 5.2. I account for these issues by using stringent fixed effects: attack level and fund family level.

While the voting data is available at the individual mutual fund portfolio level, I analyze voting outcomes at the attack × fund family level. The voting outcome at the fund family level is more reasonable as an overwhelming fraction of fund families coordinate the votes across their funds Ashraf et al. (2012); Morgan et al. (2011). Thus, using a fund family level outcome is more in line with the independent and identically distributed assumption on errors (Bolton, Li, Ravina, and Rosenthal 2020). Moreover, instead of predicting one electoral outcome for an attack, I furcate the voting at the fund family level. Thus, the predicted attack text's alignment with fund families is correlated across an attack. The correlation could reduce the standard error of β and boost significance. Clustering at the attack level mitigates this risk. I also try robust standard errors (unreported), and the results are similarly significant.

My training sample, which contains proposals at the annual meeting, is not the same as my prediction sample, the attack text. The discrepancy occurs because of the shortage of attacks that reached a voting stage. In Section 5.2, I mention that only 199 attacks went for voting over the 2004–2019 period, which is not enough to run a machine learning algorithm. To mitigate the differences between training and prediction sample, I filter out all the management proposal and train the SVR on shareholder proposals only. Shareholder proposals are often more in line with activist's proposals in attack text. Moreover, over the 2003–2018 period, fund families' voting on shareholder proposals (44% against the management) is in line with the voting in attack proposals (48% against the management).

I focus on attacks to illustrate the importance of persuasion in fund family voting. However, attack proposals make a small portion of all the voting decisions. In any year, less than twenty attacks reach the voting stage, while prominent mutual fund institutions, on average,

cast over 30,000 votes at American public companies (NYTimes 2019). Nonetheless, attacks have higher stakes for all parties involved compared to routine proxy votings for which investor votes are mostly precatory (Brav, Jiang, Li, and Pinnington 2018; Buchanan, Netter, Poulsen, and Yang 2012; Klein and Zur 2009). The attack texts are often more informative, to the point, and without boilerplate or legal jargon. Thus, by analyzing voting in attacks, I provide evidence of activists successfully persuading fund families in events that have a long-term effect on the economy.

Lastly, I do not differentiate between various channels, which leads activists to tailor their campaigns. First, the fund families maybe dissatisfied with operations of firms their portfolio, and approach activists to target these firms to elicit changes. Second, activists decide the issues they want to raise and then choose a target that has investment from fund families that are interested in those issues. Third, activists know the preferences of the shareholders and accordingly use language to persuade the shareholders. While all three channels support my results that ownership structures influence campaigns, my analysis does not identify the channels.

F Additional information for robustness tests

F.1 Classification of proposals into types

Table 15:
Shareholder proposal classified into types
This table classifies shareholder proposals into 25 proposal types based on their description in the ISS database. I start grouping proposals together beginning from the most frequent description; as such, the 25 types listed below cover 90% of shareholder proposals over the 2003–2018 period.

Prop. type	General description of proposals in ISS database	# of sh. prop.
1	Elect Directors (Opposition Slate); Elect a Shareholder-Nominee to the Board (Proxy Access Nominee); Elect Director (Cumulative Voting or More Nominees Than Board Seats).; Elect a Shareholder-Nominee to the Board; Elect Director Nominated by Preferred Shareholders; Elect Directors (Bundled Dissident Slate)	1918
2	Require Independent Board Chairman	663
3	Declassify the Board of Directors	629
4	Political Contributions Disclosure	530
5	Require a Majority Vote for the Election of Directors	498
6	Company-Specific – Shareholder Miscellaneous	358
7	Amend Articles/Bylaws/Charter – Call Special Meetings	301
8	Advisory Vote to Ratify Named Executive Officers'Compensation	289

Continued on next page

Table 15 – continued from previous page

Prop. type	General description of proposals in ISS database	# of sh. prop.
9	Company Specific-Governance Related; Company-Specific Board-Related; Amend Articles/Bylaws/Charter – Non-Routine.; Approve Recapitalization Plan for all Stock to Have One-vote per Share; Eliminate or Restrict Severance Agreements (Change-in-Control); Amend Vote Requirements to Amend Articles/Bylaws/Charter; Establish Other Governance Board Committee; Adopt Proxy Access Right; Establish Environmental/Social Issue Board Committee; Require Director Nominee Qualifications (Excluding Environmental & Social); Submit SERP to Shareholder Vote; Change Size of Board of Directors; Establish Term Limits for Directors; Proxy Voting Tabulation; Amend Proxy Access Right; Approve/Amend Terms of Existing Poison Pill; Require More Director Nominations Than Open Seats; Require Majority of Independent Directors on Board; Amend Articles/Bylaws/Charter to Remove Antitakeover Provisions; Elect Supervisory Board Members (Bundled).; Elect a Shareholder-Nominee to the Supervisory Board.; Proxy Voting Disclosure, Confidentiality, and Tabulation; Amend articles/bylaws/charter – Filling Vacancies; Require Environmental/Social Issue Qualifications for Director Nominees; Adopt Policy on Succession Planning; Amend Articles Board-Related; Establish SERP Policy; Reimburse Proxy Contest Expenses; Limit Composition of Committee(s) to Independent Directors; Establish Director Stock Ownership Requirement; Provide for Confidential Voting (INACTIVE); ...	1184

Continued on next page

Table 15 – continued from previous page

Prop. type	General description of proposals in ISS database	# of sh. prop.
9	Amend Articles/Bylaws/Charter – Removal of Directors; Amend Articles/Charter Equity-Related.; Eliminate or Restrict Shareholder Rights Plan (Poison Pill); Establish a Compensation Committee; Rotate Annual Meeting Location; Establish Shareholder Advisory Committee; Proxy Voting Disclosure; Establish Mandatory Retirement Age for Directors; Establish a Nominating Committee; Restore Preemptive Rights of Shareholders (INACTIVE)	1184
10	Restore or Provide for Cumulative Voting	245
11	Proxy Access	232
12	Submit Shareholder Rights Plan (Poison Pill) to Shareholder Vote	210
13	Stock Retention/Holding Period; Double Trigger on Equity Plans; Compensation- Miscellaneous Company Specific; Limit/Prohibit Executive Stock-Based Awards; Review Executive Compensation (INACTIVE); Pay For Superior Performance; Report on Pay Disparity; Clawback of Incentive Payments; Miscellaneous – Equity Related; Expense Stock Options (INACTIVE); Limit Executive Compensation; Link Executive Pay to Social Criteria; Increase Disclosure of Executive Compensation; Death Benefits/Golden Coffins; Non-Employee Director Compensation; Disclose Information on Compensation Consultant; Put Repricing of Stock Options to Shareholder Vote; Adjust Executive Compensation Metrics for Share Buybacks	1214
14	Remove Existing Directors	202
15	Political Lobbying Disclosure	200

Continued on next page

Table 15 – continued from previous page

Prop. type	General description of proposals in ISS database	# of sh. prop.
16	Provide Right to Act by Written Consent	199
17	Social Proposal	196
18	Improve Human Rights Standards or Policies	189
19	Performance-Based and/or Time-Based Equity Awards	188
20	Reduce Supermajority Vote Requirement	173
21	Report on Sustainability; GHG Emissions; Climate Change; Community-Environmental Impact; Report on Climate Change; Animal Welfare; Renewable Energy; Report on Environmental Policies; Recycling; Nuclear Power - Related; Environmental - Related Miscellaneous (INACTIVE); Energy Efficiency; Toxic Emissions; Toxic Substances (INACTIVE)	793
22	Appoint Alternate Internal Statutory Auditor(s) [and Approve Auditor's/Auditors'Remuneration].; Limit Auditor from Providing Non-Audit Services; Auditor Rotation; Appoint Internal Statutory Auditor(s) Nominated by Preferred Shareholders [and Approve Auditor's/Auditors' Remuneration]	85
23	Adopt Sexual Orientation Anti-bias Policy	106
24	Submit Severance Agreement (Change-in-Control) to Shareholder Vote	92
25	Board Diversity; Report on EEO	125

F.2 Confidence interval for varying parameters

Figure 12:
The positive association between attack's aggregate alignment and activist's success holds for changing ownership dummy cutoff.
This figure plots β coefficient with 95% confidence interval for regression of attack outcome on fund family holdings weighted attack text'alignment. Specifically, I estimate:

$$Win_a = \gamma AgAlign_a + \lambda OwnDum_a + \beta AgAlign_a \times OwnDum_a + \epsilon_a$$

where Win_a represents a dummy, which is one if the result of the attack, *a*, is Successful or Settled. *AgAlign* is the holdings-weighted attack text'alignment with fund families. *OwnDum* is the ownership dummy that is one if the mutual funds, whose voting information is available, own more than the cutoff of target shares. The sample consists of attacks that went to a voting stage over the 2004–2019 period. The independent variable, *AgAlign*, is scaled by the standard deviation of the underlying variable. The standard errors, ϵ_a, is robust and computed with the sandwich estimator of variance.

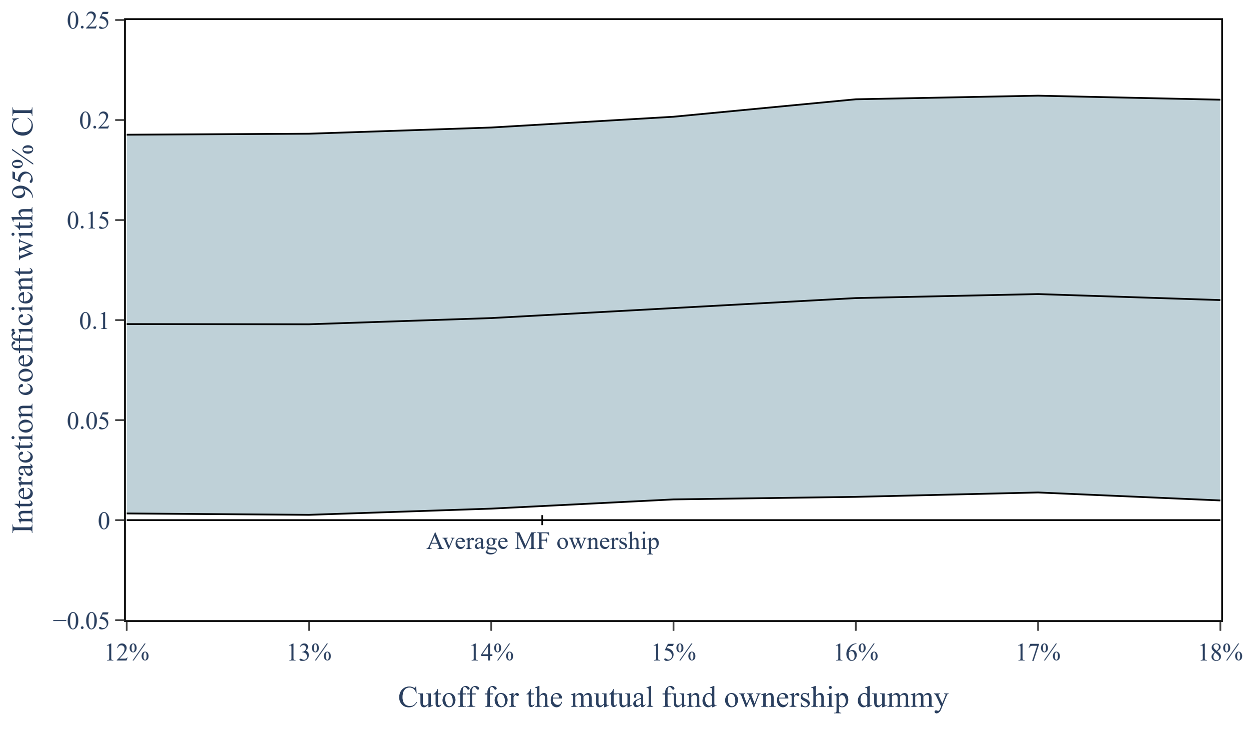
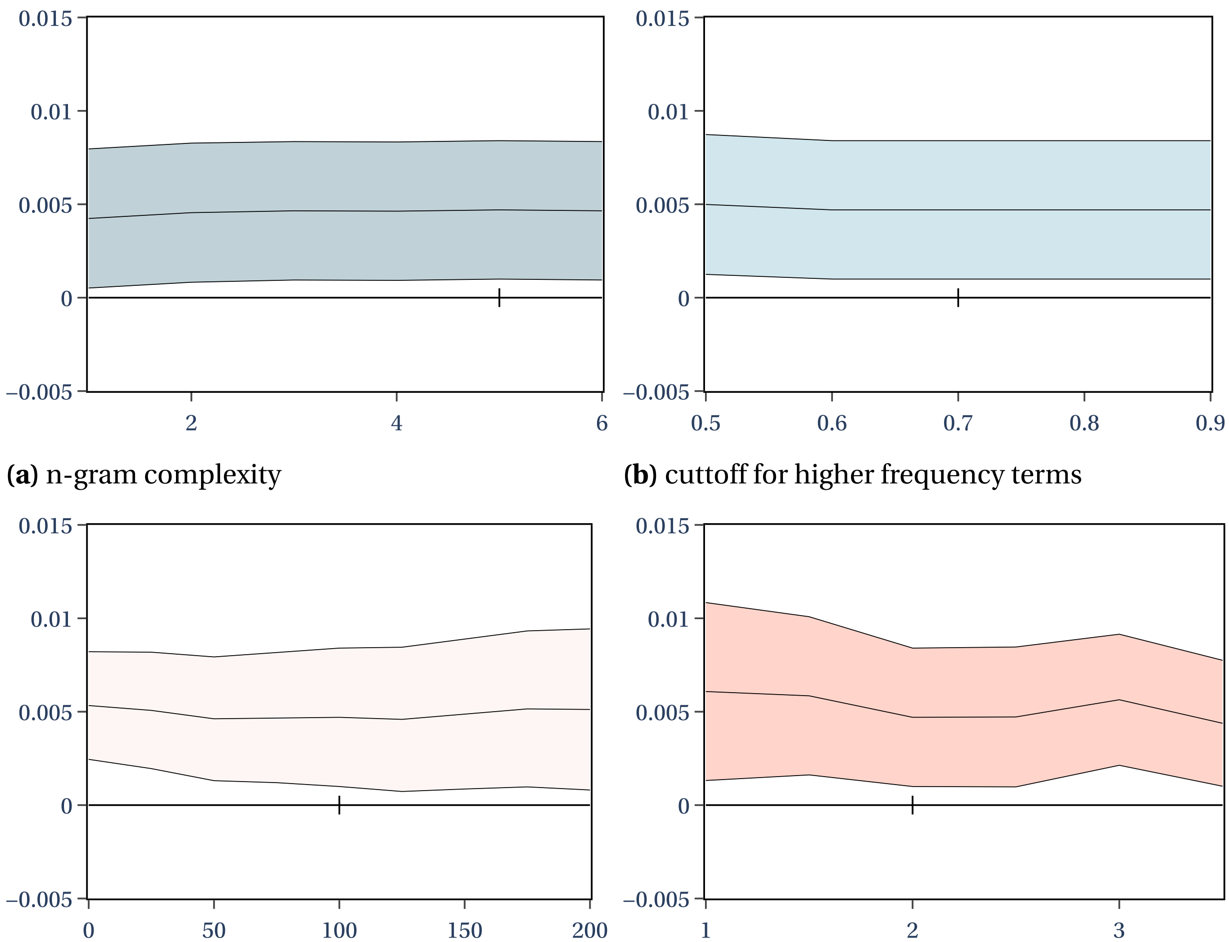


Figure 13:
The positive association between fund family holdings and attack text's alignment is robust to changing SVR parameters.
This figure plots β coefficient with 95% confidence interval for regression of fund family's text-based likelihood of supporting activists on fund family holdings in the target. Specifically, I estimate:

$$\widehat{Align}_{a,f} = \beta Holding_{a,f} + \delta_a + \delta_f + \epsilon_{f,a}$$

where $\widehat{Align}_{a,f}$ is the predicted alignment of attack text, *a*, with the fund family, *f*, preferences. *Holding* is the percent of equity the fund family owns of the target before the attack, obtained from CRSP database. δ_a, and δ_f represent attack level and fund family level fixed effects, respectively. The sample consists of attacks, identified using SEC filings, over the 2004–2019 period. Corresponding fund families include all the families that have voted in at least a hundred shareholder proposals in the two years prior to the attack. The independent variable is scaled by the standard deviation of the underlying variable, meaning the coefficient can be interpreted as the effect of a one-standard-deviation change in the determinant. Standard errors, $\epsilon_{f,a}$, are clustered at the attack level.



(a) n-gram complexity

(b) cuttoff for higher frequency terms

(c) minimum shareholder proposals required

(d) past voting considered (in years)

Bonds Lie in the Portfolio of the Beholder: Do Bonds Affect Equity Monitoring? ☆

Todd A. Gormley[a,*] and Manish Jha[a]

[a] *Washington University in St. Louis, Olin Business School, St. Louis MO 63130, USA*

Click here for the most recent version[ǂ]

November 10, 2020

ABSTRACT

We analyze whether institutional investors' increasingly extensive corporate bond holdings are associated with how actively institutions vote and monitor their equity investments. We find that institutions conduct more governance research and are less likely to follow proxy advisor vote recommendations for companies whose bonds represent a larger proportion of their overall portfolio. Corporate bonds held in equity-focused funds and shareholder proposals that are more likely to require investors' attention drive these findings. There is no evidence that creditor-shareholder conflicts explain these findings. Our results suggest that institutions' bond holdings contribute to their overall incentive to be engaged monitors.

JEL classification*:* G23, G30, G32, G34, K22
Keywords*:* bonds, governance, institutional investors, attention, voting

☆ For helpful comments, we thank the seminar participants at the "early ideas" session of the Financial Research Association conference.

[*] Corresponding author. Tel: +1 (314) 935-7171
E-mail address: gormley@wustl.edu (Todd A. Gormley).

[ǂ] Article's web link: *https://ssrn.com/abstract_id=3719187*

"[I]t can be challenging for investors to consider how to adopt their stewardship practices to include fixed income… Yet in many areas of corporate governance, there can be a significant alignment of interest that supports engagement on behalf of all financial stakeholders, both creditors and shareholders."
— George S. Dallas, Policy Director at International Corporate Governance Network

1. Introduction

The increasingly diverse holdings of institutional investors, who hold around 70% of public US firms' outstanding equity, raise questions about whether they actively vote their shares and monitor individual companies. For example, each of the three largest mutual fund families in terms of total net assets (Vanguard, Fidelity, and BlackRock) held equity positions in around 5,000 US companies as of December 2018, casting doubt on their ability to monitor every company in their extensive portfolios. However, recent evidence suggests these institutions are attentive owners for stocks that comprise a larger portion of their overall portfolio (Fich, Harford, and Tran, 2015; Iliev and Lowry, 2015) and that even institutions primarily holding indexed positions play an important governance role (e.g., Appel, Gormley, and Keim, 2016). This paper analyzes whether institutions' large *corporate bond* portfolios might affect how actively they monitor and vote their shares.

Institutions typically offer various mutual fund and ETF options to investors (e.g., equity-only, bond-only, and mixed asset), and corporate bond holdings comprise an increasingly significant component of many institutions' portfolios. The average share of a mutual fund family's total net assets (TNA) held in corporate bonds has increased from around 5% in 2008 to 15% in recent years (see Figure 1). Moreover, 34.1% of institutions casting votes on contentious shareholder proposals between 2008 and 2018 also held a bond position in the underlying firm; and the bond position accounted for, on average, 28.1% of the institution's exposure to the firm.

There are various reasons to think these bond holdings could affect how actively institutions vote and monitor their equity positions. First, fund managers typically research a company and its default risks when they hold a company's bonds. If shared internally within the larger fund family, this additional information could affect how informed and attentive the

institution is when voting their shares.[1] Because credit rating agencies increasingly incorporate companies' activities on environmental, social, and governance (ESG) issues into their assessments of firms' risk, fund managers with a bond position also have incentives to encourage active voting and ownership by their equity counterparts in the larger fund family.[2] Additionally, relative to equity fund managers who might expect to only hold their position for a couple of years, managers holding a bond position might be more attentive to company-specific events essential to the viability of a company and its ability to meet long-term debt obligations (Dallas, 2019).[3]

On the other hand, there are also reasons to believe that an institution's corporate bond portfolio will not affect how attentive it is to a particular company or how actively it votes on any given proposal. Unlike equity owners, bond investors have no voting rights, and therefore, might have little input into how equity fund managers within the larger fund family vote their positions. Additionally, concerns about violating their fiduciary duty to act on behalf of their investors could limit the extent to which equity fund managers allow other managers with a bond position (and hence, a potentially conflicting interest) to influence their voting decisions.

To analyze the importance of bond holdings for investor attention, we employ a panel estimation to determine whether the size of an institution's bond position is associated with a proxy for investor attention. We construct a proposal-by-institution-level dataset of how institutions voted on every proposal from 2008 to 2018 and pair this data to institutions' aggregated holdings at the time of the vote. We then regress a proxy for how actively an institution voted on a shareholder proposal onto the share of the fund family's TNA held in that company's bonds.

---

[1] This information spillover could occur in a variety of ways. For example, many large institutions centralize voting decisions in governance divisions that aggregate fund managers' views and information before casting votes. Interviews confirm that individuals in these governance divisions consult both bond and equity fund managers. In institutions where individual fund managers make voting decisions, it is possible that equity managers seek the input of managers with bond positions before voting. Analyzing investment decisions within fund families, Auh and Bai (2020) find evidence consistent with cross-asset information spillovers.

[2] In April 2019, Moody's acquired a majority stake in Vigeo Eiris, a global leader in ESG research, data, and assessments, as part of its ongoing push to incorporate such information into credit ratings. S&P Global also uses ESG assessments in their credit rating process (Dallas, 2019).

[3] Consistent with this, the lead governance director at a large institution described managers with bond positions as more "paranoid and pessimistic" than equity-only managers. Thus, their presence within the larger fund family resulted in added pressure for his governance division to monitor firms closely.

Following Iliev and Lowry (2015) and Gilje, Gormley, and Levit (2020), we proxy for an institution's attention by examining whether its votes went against the recommendations of the proxy advisory firm Institutional Shareholder Services (ISS). The underlying premise is, all else equal, attentive investors are less likely to rubber-stamp ISS recommendations.

To mitigate concerns about portfolio weights' endogeneity, we partial out potential confounding factors that might drive differences in attention at the investor- or proposal-level. In particular, we include proposal-level fixed effects in each estimation. Their inclusion accounts for any firm- or proposal-level characteristics that affect institutions' likelihood of voting against ISS recommendations and allows us to isolate how votes on a given proposal vary across institutions as a function of their bond holdings. We also include institution-by-month fixed effects to control for each institution's general tendency to vote against ISS and the possibility this might vary over time. In other words, we only use variation in how an institution voted across proposals in a given month as a function of how extensive its bond position was in each company it voted on.

Using this within-proposal and within-institution-by-month variation in votes, we find a positive association between the size of an institution's bond position and the likelihood it does not follow the ISS recommendation when voting its shares. The positive association is robust to controlling for the overall importance of institutions' equity position as a share of TNA, which prior work finds is positively associated with being attentive (e.g., Fich, Harford, and Tran, 2015; Iliev and Lowry, 2015; Gilje, Gormley, and Levit, 2020). The magnitude of the association between bond holdings and voting is economically important. A one standard deviation increase in a bond's share of TNA is associated with a shift in voting that is one-third of the observed change for a one standard deviation increase in an equity's share of TNA.

The observed association between bond holdings and voting varies across funds and proposals in ways consistent with a potential influence on investor attention. During our sample period, 17% of funds classified by Lipper as equity-focused also contain corporate bonds, and for those funds, bonds account for about 21% of holdings. Consistent with managers of equity-focused funds being more likely to influence an institution's voting decisions, we find that bonds held in

such mixed-asset funds matter more for our findings. We also find that proposals more likely to require investor attention, including contested proposals (where the final vote margin was within five percentage points of the level needed for passage) and contentious proposals (where ISS recommended voting against management), drive our findings.

We next construct an alternative proxy for investor attention, the number of times a fund family accesses a company's SEC filings via EDGAR in the days before a shareholder meeting. Following Iliev, Kalodimos, and Lowry (2020), we construct this measure of governance research by matching the IP addresses accessing each filing on EDGAR to individual fund families using a linking table that records the registered owner of each IP address. To proxy for investor attention, we aggregate each institution's number of views from 30 days before releasing the meeting's proxy statement and continuing through the shareholder meeting date. Because this alternative measure is only available at the meeting level, we replace our proposal-level fixed effects with meeting-level fixed effects but retain the institution-by-month fixed effects.

The size of an institution's bond holdings is positively associated with the number of times the institution views a company's filings before a shareholder meeting. The association is robust to controlling for the size of the institution's equity position and of similar magnitude to what we observe using our first proxy of investor attention. One standard deviation higher bonds as a share of TNA is associated with increased EDGAR viewings of that company's filings that is about 40% of the observed magnitude for one standard deviation higher equity as a share of TNA. Like our first proxy for investor attention, meetings with a contested or contentious shareholder proposal and bond positions held in equity-focused funds drive these findings.

Next, we analyze whether the observed pattern varies based on fund type or institution type. We find that the positive association between portfolio weights and investor attention is only present for bonds held in actively managed equity funds. Bond positions held in index funds have no association with investor attention. However, when we separately analyze the importance of bonds for "The Big Three" indexers (BlackRock, State Street, and Vanguard), which account for a combined 75% of all indexed assets, we find that even indexed bond holdings are associated with

increased attention for these institutions. We also find that the importance of a position's portfolio weight is more significant for The Big Three. The latter finding is consistent with these institutions being especially likely to allocate their limited attention to their most significant holdings.

Shareholder-creditor conflicts arising from institutions' dual holdings do not drive our findings. The association between institutions' bond holdings and voting patterns is mostly unchanged when excluding firms in financial distress, where a wedge in the interests of shareholders and creditors is more likely. Nor do we find evidence the association is more significant for firms in financial distress. Shareholder-creditor conflicts also do not explain our findings regarding institutions' governance research, as measured using EDGAR viewings.

These findings have important implications for corporate governance and the monitoring of managers. Institutional investors are not fully attentive (e.g., Ben-Rephael, Da, and Israelsen, 2017; Fang, Peress, and Zheng, 2014; Lu, Ray, and Teo, 2016; Schmidt, 2019), particularly when it comes to voting and being engaged monitors of their smaller equity positions (Fich, Harford, and Tran, 2015; Iliev and Lowry, 2015; Iliev, Kalodimos, and Lowry, 2020). This lack of attention affects managers' incentives and destroys shareholder value (e.g., Kempf, Manconi, and Spalt, 2017; Liu, Low, Masulis, and Zhang, 2020; Gilje, Gormley, and Levit, 2020). However, our findings show that institutions with extensive corporate bond holdings are more attentive, suggesting the growing popularity of mixed asset funds and institutions' tendency to hold both a bond and equity position in companies can enhance investor stewardship.

Our findings also contribute to the nascent literature that quantifies institutions' incentive to be engaged monitors. Existing estimates of institutions' motive to monitor consider how improvements in their equity positions' value will increase their fund fees and flows (Lewellen and Lewellen, 2020). However, this focus on institutions' equity positions ignores that active monitoring can also improve the value of institutions' bond holdings, providing many institutions an additional motive to be engaged owners. Our findings suggest that institutions' combined debt and equity holdings and the type of funds invested in those positions should be accounted for when proxying for institutions' overall incentive to be engaged owners.

Finally, our results contribute to the ongoing discussion regarding the conflicting interests of institutions that hold both debt and equity. Debt and equity owners can have different views regarding the value implications of dividends, equity issuances, takeover defenses, and acquisitions, which could influence how an institution that holds both debt and equity votes on particular proposals. Consistent with this, evidence suggests that institutions holding both debt and equity in a firm vote differently on proposed mergers (Bodnaruk and Rossi, 2016) and are more likely to cast creditor-friendly votes when the firm is in financial distress (Keswani, Tran, and Volpin, 2020). We instead analyze the importance of debt holdings for an institution's overall likelihood of being an engaged monitor, which can positively influence the value of both debt and equity positions. We find evidence that these dual holdings influence institutions' stewardship more generally and in ways that do not necessarily forgo equity investors' interests.[4]

We organize the paper as follows. Section 2 describes our data. Section 3 presents our empirical specification, and Section 4 reports our main findings. Section 5 analyzes heterogeneity in the importance of bonds across funds and institutions, and Section 6 examines whether our findings differ for firms in financial distress. Section 7 concludes.

2. Data and summary statistics

To assess the association between an institution's bond holdings and its level of attention to individual companies, we combine various datasets, including mutual funds' holdings, mutual fund voting records, and SEC log files of institutions accessing proxy filings.

2.1. Mutual fund holdings data

To calculate how significant each company's equity or bonds are in a fund family's overall portfolio, we use the CRSP Mutual Funds Database. Mutual funds and exchange-traded funds (ETFs) are required by the SEC to disclose their holdings quarterly during their fiscal year using

[4] In this regard, our findings also differ from papers that use the dual debt and equity holdings of banking institutions and investors to study the effects of shareholder-creditor conflicts on investment, risk-shifting, loan spreads, the use of debt covenants, and the resolution of financial distress (e.g., Chava, Wang, and Zou, 2019; Chu, 2018; Chu, Diep-Nguyen, Wang, Wang, and Wang, 2020; Jiang, Li, and Shao, 2010; Yang, 2019).

Forms N-CSR and N-Q. Many funds, however, voluntarily report holdings on other dates as well.[5] We restrict our analysis to holdings starting in 2008 because the CRSP database contains inaccurate information before that year (Schwarz and Potter, 2016).[6]

To analyze how holdings correlate with subsequent institution-level measures of attention, we aggregate security holdings to the institution (i.e., fund family) level for each month. To construct this monthly measure, we aggregate across all of the most recent fund reports of a particular institution going back three months. Because funds are required to report quarterly, the 3-month window captures the holdings of each fund.

To aggregate holdings to the institution level, we manually match funds to the larger fund family using their name while accounting for subsidiaries within each institution. For example, Allianz purchased both Nicholas-Applegate Capital Management and Pacific Investment Management Company (PIMCO) in 2000, and in 2008, it invested $2.5 billion in Hartford Financial Services Group. Because our sample begins in 2008, we assign all funds with names containing "Allianz," "Nicholas-Applegate," "PIMCO," and "Hartford" to the Allianz fund family. When aggregating to the institution level, we exclude positions with a negative value. Our findings are similar if we instead keep these negative positions or use their absolute value when aggregating. Finally, we use WRDS's CUSIP-PERMCO link table to assign a PERMCO to each security in our sample, where each PERMCO identifies a unique firm.

Because the CRSP database does not directly flag whether reported securities are a bond, we classify securities as a "bond" using two methods. First, we classify securities that report a value in the "Date of Bond Maturity" field as bonds. Because this field is missing for some bonds,

---

[5] Because most funds' fiscal year aligns to the calendar year, mandated disclosures typically occur on the last days of March, June, September, and December each year. However, some funds also make additional voluntary disclosures to significant databases, like CRSP, Morningstar, and Thomson Reuters, on other dates. However, most of these voluntary disclosures also occur at the end of March, June, September, and December (Gormley, Kaplan, and Verma, 2020). Institutions already having to disclose their aggregated holdings to the SEC on Form 13F on these same dates likely drives these additional disclosures (Schwarz and Potter, 2016).

[6] In 2008, CRSP migrated to using Lipper data instead of Morningstar data, which seems to have resulted in an increase in its coverage of SEC-mandated disclosures (Schwarz and Potter, 2016). We use the CRSP mutual fund holdings rather than the other commonly used dataset for such holdings, Thomson Mutual Fund Holdings (Thomson), as it is easier to merge to voting outcomes, resulting in a smaller loss of observations. Moreover, Schwarz and Potter (2016) document that the CRSP Mutual Fund Database has better coverage than Thomson after 2007.

we also flag a security as a bond if the security's name includes a "%", ".", "-", "/", or any number. These symbols and numbers appear in a security name for bonds to indicate a maturity date and yield rate. For example: "RAYTHEON CO., 7.20%, 8-15-2027" has a blank maturity date in the CRSP database but refers to Raytheon's 7.2% domestic bond expiring in 2027. We classify all other securities as "equity," and a manual review of the resulting security classifications confirms that this approach accurately flags bond and equity securities.

Bond holdings comprise a sizable and growing component of institutions' portfolios. Figure 1 plots the weighted average share of total assets held in bonds between 2008 and 2018 for all fund families and the three largest fund families in terms of total net assets (Vanguard, BlackRock, and Fidelity). In 2008, bonds accounted for about 5% of mutual fund and ETF assets, but this average increased to about 15% by 2018. Figure 2 provides a breakdown of bond holdings across fund families. At the end of 2018, mutual fund families held about $2 trillion in bonds. Of this, Vanguard held $552 billion, while BlackRock held $217 billion.

There is considerable variation in the importance of corporate bond holdings across fund families. Table 1, which provides a breakdown between equity and bonds for some of the largest mutual fund families (after excluding institutions' government bond holdings), shows this variation. Of The Big Three, BlackRock holds the largest share of its assets in corporate bonds, 19.2%, while State Street has the smallest share, 5.4%. Corporate bonds account for 15.7% of Vanguard's assets. There is also variation in bond holdings across the largest institutions that primarily offer actively managed funds. More than half of Allianz's $225 billion in assets is in corporate bonds, compared to just 7.8% of T Rowe Price's $397 billion in assets.

Fund-level summary statistics also show the importance of bonds for mutual fund families. Most mutual fund families offer a range of funds, including equity-only, bond-only, and mixed-asset mutual funds. Table 2, Columns (1)-(3) provides a yearly breakdown of such funds. While bond-only funds account for about 2% of all funds in 2008 and 2009, they have grown in popularity since, accounting for 11% of all funds by the end of 2018. Mixed-asset funds, which hold both corporate debt and equity securities, are also relatively common, accounting for 16–25% of all

mutual funds and ETFs per year between 2008 and 2018. These mixed-asset funds hold, on average, 45% of their assets in corporate bonds (Table 2, Column 4).

In our later analysis, we also look at how an institution's attention varies based on the type of fund that holds the bonds. Specifically, we test whether the association between bond holdings and fund families' level of attention varies when the bonds are held in an equity- versus bond-focused fund. To assign funds as an "equity" or "bond" fund, we rely on the Lipper asset classifications provided in CRSP. There are three mutually exclusive Lipper classifications: Equity Funds (EQ), Taxable Fixed Income Funds (TX), and Tax-Free Fixed Income Funds (MB). We classify the former as "equity" funds and the latter two as "bond" funds.

Even equity-focused funds have substantial bond components. As shown in Table 2, Column (5), Lipper classifies between 44% and 67% of mixed-asset funds as equity funds. Among all equity-focused funds, about 17% hold some bonds, and of those, 21% of their overall assets, on average, are held in bonds (see Table 2, Columns 6-7).[7] A similar pattern holds for bond-focused funds. In a given year, between 46% and 87% of bond funds contain an equity component that accounts for, on average, about 17% of holdings (Columns 8-9).

2.2. Mutual fund voting data

We use the ISS Voting Analytics dataset to analyze how institutions' votes vary as a function of their bond holdings. The database includes fund voting records, as obtained from the mandated N-PX forms that institutions file with the SEC every year. While the voting records are available from July 2003 to October 2018, we start our sample in 2008 to match the time for which we have fund holdings data and to match when the coverage of Voting Analytics is better. Before 2007, ISS only collected voting records of the top 100 fund families, but after 2007, it collected the top 300 fund families (Brav, Jiang, Li, and Pinnington, 2019). In addition to recording each fund's votes, the ISS data also includes a description of the proposal, the ISS recommendation on

[7] Some example funds classified as equity funds by Lipper are the Fidelity Balanced Fund (where bonds account for about 27% of assets), BlackRock's Multi-Asset Income Fund (where bonds are 37% of assets), Vanguard's Wellington Fund (where bonds are 33% of assets), T Rowe's Capital Appreciation Fund (where bonds are 22% of assets), and the Hartford Balanced Income Fund of Allianz (where bonds are 48% of assets).

how investors should vote, the threshold required for passage, and the final vote outcome.

For our baseline analysis, we follow Iliev and Lowry (2015) and Gilje, Gormley, and Levit (2020) and focus on shareholder-sponsored proposals. During our sample, there are 9,331 proposals sponsored by shareholders, and of these, 4,972 (or 51.1%) are contentious, as defined by when ISS and management gave conflicting vote recommendations. For our main tests, we exclude non-contentious proposals because they are typically not well-thought-out (Gantchev and Giannetti, 2019) and because investors do not appear to focus on them (Iliev, Kalodimos, and Lowry, 2020). A similar logic applies to excluding management proposals, which are mostly perfunctory and less revealing about investor attention (Iliev and Lowry, 2015; Gilje, Gormley, and Levit, 2020). However, in subsequent tests, we show how our findings differ across the various proposal types, including management and non-contentious proposals.

We aggregate the fund-level votes to the institution (i.e., fund family) level using the same approach used to aggregate mutual fund holdings and then merge the voting data to the holdings data. When merging in the holdings data for each proposal-by-institution observation, we use the aggregated holdings across all the most recent fund reports of the institution in the three months before the proposal is voted on. After this merger, we have 369 unique institutions in our sample, and, on average, 56 institutions and their funds cast votes for each proposal. In total, our sample includes 276,566 proposal-by-institution observations across 11 years.

Because investor attention is unobservable, we follow Iliev and Lowry (2015) and proxy for it using an indicator for whether an institution's votes on a proposal fail to follow the ISS recommendations. Iliev and Lowry (2015) and Malenko and Malenko (2019) posit that if fund families devote more resources towards becoming informed, they will be less likely to follow proxy advisory firm recommendations indiscriminately. Consistent with this, Iliev and Lowry (2015) observe a greater likelihood of disagreeing with ISS for mutual funds where the net benefits of being attentive are greater. Moreover, Iliev, Kalodimos, and Lowry (2020) find that this voting behavior is positively related to an institutional investor becoming informed before a vote.

To create our proposal-by-institution indicator for failing to follow an ISS recommendation, we follow Gilje, Gormley, and Levit (2020). We code vote decisions of

"Against," "Abstain," and "Withhold" as "Against," and "For" as "For." We then compare how each institution voted on a proposal to the ISS recommendation of either "For" or "Against" and flag those where the institution did not follow ISS. In a small number of cases where not all funds within the institution vote in the same direction (which occurs in 15.7% of cases), we use the percent of funds within the institution that chose not to follow ISS as our investor attention proxy.

For the contentious shareholder proposals we analyze, there is considerable variation across institutions on whether they follow ISS. Table 3 provides summary statistics for our final proposal-by-institution sample. On average, 47.3% of institutions cast a vote that does not follow the ISS recommendation, and 40.2% of funds cast a vote that does not agree with ISS. While not tabulated, we find that the average likelihood of voting against ISS is considerably lower for management proposals (7.6%) and non-contentious shareholder proposals (9.8%), consistent with these excluded proposals being routine and less likely to require investors' attention.

The summary statistics provided in Table 3 further highlight the potential importance of institutions' bond holdings. In 34.1% of observations, the voting institution held a non-zero bond position in the company. On average, these bond holdings accounted for about 0.035% of the institution's overall total net assets and 28.2% of their overall exposure to the firm.

In many cases, a voting institution also holds bonds in the company as part of an equity-focused fund. Table 3, where we break bond holdings into those found in bond-focused (i.e., Bond Holdings [in Bond Funds]) and equity-focused funds (i.e., Bond Holdings [in Equity Funds]), illustrates this. In 31.4% of observations, the voting institution holds a non-zero bond position in one of its bond-focused funds. In 16.0% of observations, the voting institution holds a non-zero bond position in one of its equity-focused funds. In Section 4.2, we analyze whether the type of fund holding the bonds is associated with investor attention.

2.3. Mutual fund's accessing of company filings on EDGAR

As an additional proxy of investor attention, we use how often an institution accessed the company's SEC filings before a meeting. Other papers have used investors accessing of SEC filings as a measure for corporate governance research (e.g., see Bauguess, Cooney, and Hanley 2018; Loughran and McDonald, 2017; Iliev, Kalodimos, and Lowry, 2020).

To measure how often an institution accessed a company's SEC filings before a meeting, we use the publicly available EDGAR log files. SEC's Division of Economic and Risk Analysis (DERA) assembles information on internet search traffic for EDGAR filings through SEC.gov, covering February 14, 2003, through June 30, 2017. The log file contains the first three octets of the IP address accessing each file and a time stamp on when the file was accessed. To assign these IPs to institutional investors, we use a linking table purchased from Digital Elements, an IP geolocational technology provider, containing names of the organizations registering each IP address as of December 31, 2016. We then follow the approach recommended by Iliev, Kalodimos, and Lowry (2020) to match these organization names to specific institutional investors. See Appendix A for details of this matching process.

We follow Iliev, Kalodimos, and Lowry (2020) in constructing a meeting-by-institution level of investor attention. Specifically, we count the number of times an institution accessed any document filed by the firm in the days before a meeting, so long as at least one of these was a proxy document associated with the upcoming meeting.[8] We identify proxy documents using the accession number provided by SEC index files. Our count aggregates the number of daily views from 30 days before the proxy statement date and continuing through the shareholder meeting date. Typically, proxy statements are released 45 days before the shareholder meeting, resulting in an average window of 75 days. We follow Iliev, Kalodimos, and Lowry (2020) and use the log of this count to mitigate the skewness in count numbers. We also show the robustness of our findings to instead using an indicator for a non-zero number of viewings.

During our sample period, January 2008 to June 2017, we obtain log files for 41,996 shareholder meetings and can identify 141 unique institutions. After limiting our sample to institutions with non-zero equity holdings on the meeting date, our final sample includes 1.29 million institution-by-meeting observations. On average, 31 institutions download proxy filings before a shareholder meeting. The mean number of views is only 0.11, as 91.8% of the institutions

[8] Iliev, Kalodimos, and Lowry (2018) also propose a narrower count that only includes the number of times the institution accesses the proxy statements associated with the meeting. Our subsequent findings are robust to using this alternative measure of investor attention instead.

with an equity position do not access the proxy filing before the meeting. However, when an institution does download filings, the average stands at 2.64 views.

3. Estimation strategy

To analyze the association between an institution's level of attention and the importance of a particular bond position in an institution's overall portfolio, we start by estimating

$$Against_{ijkm} = \beta(\frac{Bond}{TNA})_{ikm} + \gamma(\frac{Equity}{TNA})_{ikm} + \alpha_j + \delta_{im} + \varepsilon_{ijkm}, \quad (1)$$

where *Against* is an indicator for institution i voting against the ISS recommendation on proposal j for firm k in month m, *Bond/TNA* is the proportion of institution i's total net assets (TNA) held in firm k's bonds as of month m, *Equity/TNA* is the proportion of institution i's TNA held in firm k's equity, and α_j and δ_{im} are proposal and institution-by-month fixed effects, respectively. To ensure outliers do not unduly influence our findings, we winsorize both *Bond/TNA* and *Equity/TNA* at the one percent level. Furthermore, to ease the estimates' interpretation, we scale both *Bond/TNA* and *Equity/TNA* (and subsequent explanatory variables) by their sample standard deviation. Thus, each variable's coefficient reflects the change in the outcome for a one standard deviation increase in that variable. Because the estimation errors, ε, might exhibit serial correlation and be correlated within institutions, we cluster the standard errors at the institution level.

Our main identification concern is that of omitted variables. If *Bond/TNA* correlates with proposal-, firm-, or institution-level characteristics that affect an institution's likelihood of actively voting its shares (i.e., not blindly following the ISS recommendation), then our estimate of interest, β, could reflect these omitted variables rather than an effect of bond holdings on investor attention. For example, if institutions tend to hold larger bond positions in better-run companies that are more likely win shareholder support, then *Bond/TNA* and *Against* could be positively correlated even if bond holdings have no effect on institutions' monitoring.

However, the inclusion of proposal and institution-by-month fixed effects allows us to control for a number of these potential omitted factors. The proposal-level fixed effects control for any proposal-level characteristics that could affect institutions' likelihood of following ISS,

including the proposal's type and content. The proposal fixed effects also control for any characteristics of the firm (e.g., profitability and size) at the time of the vote that might matter for how institutions vote on a particular proposal. The institution-by-month fixed effects control for any differences in an institution's overall tendency to be "pro-management" (e.g., Brav, Jiang, Li, and Pinnington, 2019; Kedia, Starks, and Wang, 2020), while allowing for this tendency to change over time. Hence, our coefficient of interest, β, is identified using variation in how votes for a given proposal vary as a function of each institution's bond holdings in a given month.

What these fixed effects do not control for, however, are other factors that exhibit cross-sectional variation across an institution's holdings at a particular point in time that both affect the likelihood of an institution voting against ISS and correlate with *Bond/TNA*. One such factor is how important that firm's equity is in the institutions' overall equity portfolio, which could correlate with *Bond/TNA* and affect institutions' monitoring (e.g., Fich, Harford, and Tran, 2015; Iliev and Lowry, 2015). For this reason, we also include the proportion of an institution *i*'s TNA held in firm *k*'s equity as of month *m*, *Equity/TNA*, as an additional control.[9]

Unlike papers that analyze the importance of equity holdings for investor attention, we are less concerned about a potential simultaneity bias. While an institution's desire to vote in a particular direction and influence the voting outcome of an upcoming proposal might affect an institution's desire to increase its equity position before a vote (e.g., to increase its voting power), a similar such concern would not apply to bond holdings because the size of an institution's bond holdings will not affect the number of votes it can cast.

4. Empirical Findings

This section analyzes the association between bond holdings and institutions' voting using the specification in eq. (1). We then assess how this association varies as a function of whether equity- or bond-focused funds hold the bonds, and how it varies across proposal types. We also

[9] While previous papers tend to measure the importance of an equity position relative to the overall equity portfolio, we scale an institution's equity holdings by its TNA to ensure that we are scaling institutions' bond and equity holdings in the same way. This also makes the coefficients on the two regressors directly comparable. However, our subsequent findings are robust to instead scaling *Equity* using just the total value of an institutions' overall equity portfolio.

test our findings' robustness to using the number of times an institution accesses a company's proxy statement and other SEC filings via EDGAR as an alternative proxy for investor attention.

4.1. Bond holdings and voting against ISS

To assess how bond holdings might influence institutions' voting and level of attention, we start by estimating a version of eq. (1) that excludes the *Equity/TNA* control. This estimation determines the baseline association between institutions' bond holdings in a company and institutions' likelihood of voting against ISS for that company's proposals, after controlling for proposal and institution-by-month fixed effects. Table 4, Column 1 reports the findings.

We find that institutions where the bonds of a firm represent a larger proportion of their overall portfolio are more likely to vote against the ISS recommendation on contentious shareholder-sponsored proposals for that company. Specifically, a one standard deviation increase in the share of an institution's overall portfolio held in a particular firm's bonds (0.03%) is associated with a 0.00546 percentage point increase in the likelihood of voting against ISS (Table 4, Column 1), corresponding to about a 1.2% increase relative to the sample standard deviation.

Similar to prior work analyzing how investors' attention varies with their equity holdings, the association between bond holdings and an institution's likelihood of voting against ISS is concave. To illustrate this, we follow Gilje, Gormley, and Levit (2020) and plot the point estimates from a regression of *Against* onto dummy variables for each portfolio weight quintile of *Bond/TNA*, proposal fixed effects, and institution-by-month fixed effects. Figure 3 reports the findings. Using a linear extrapolation between point estimates, we find that the association between *Bond/TNA* and *Against* is concave. The concavity indicates that the observed increase in attention for a given increase in bond holdings diminishes as the portfolio weight increases.

The positive association between bond holdings and voting is robust to controlling for the proportion of institutions' portfolio held in the firm's equity (Column 2). Consistent with the prior literature, we find a positive association between the importance of a stock in an institution's portfolio and the likelihood of that institution disagreeing with ISS (e.g., Iliev and Lowry, 2015; Gilje, Gormley, and Levit, 2020). A one standard deviation increase in *Equity/TNA* (0.43%) is

associated with a 0.0156 percentage point increase in the likelihood of voting against ISS. However, the coefficient on *Bond/TNA* remains mostly unchanged and is still statistically significant at the five percent level. After controlling for proposal- and firm-characteristics at the time of the vote (as done using the proposal fixed effects) and an institution's overall tendency to disagree with ISS (as done using institution-by-month fixed effects), institutions are more attentive voters when that firm's equity *and* bonds represent a larger proportion of the institution's portfolio.

Intuitively, the importance of an institution's bonds for its voting behavior is less than its equity holdings. Comparing the observed shift in voting for a one standard deviation change in *Bond/TNA* to a one standard deviation change in *Equity/TNA,* bond holdings are associated with about one-third of a shift in voting compared to what is observed for equity holdings. This difference makes sense. If these point estimates reflect institutions' paying more attention to their most significant holdings, we might expect it to be less for bond positions, which do not get a vote.

4.2. Heterogeneity by type of fund and proposal

Next, we assess whether the positive association between how important a firm's bonds are in an institution's overall portfolio and that institution's voting behavior depends on which type of funds hold those bonds. As shown in Table 2, an institution might have a bond position both because of holdings by its bond funds and because of holdings by its equity-focused funds that also take some debt positions, which occurs in 17% of equity-focused funds.

The type of fund that holds the bonds might matter if there are differences in the relative amount of attention paid to shareholder proposals by the managers of equity- and bond-focused funds. For example, if managers of equity-focused funds pay more attention to shareholder proposals because they must decide how to vote their positions (or what recommendation to give to their institution's proxy voting committee), then bond holdings in those funds might matter more for how attentive the fund manager (and her institution) is when voting. However, it is possible that even holdings in bond-focused funds could matter for institution-level attention if the managers of those funds volunteer their opinions or are consulted before votes.

To test for heterogeneity across fund types, we repeat our estimation of eq. (1) after

replacing *Bond/TNA* with two measures of how important a company's bonds are in the institution's portfolio. The first, *Bond [in Bond Funds]/TNA*, measures the proportion of an institution's overall net assets held in the company's bonds, where bond-focused funds hold the bonds. The second, *Bond [in Equity Funds]/TNA*, reflects the proportion of an institution's net assets held in the company's bonds, where equity-focused funds hold the bonds. The sum of these two bond measures equals the original *Bond/TNA* for a given institution-by-proposal observation by construction. In 39.1% of cases where an institution has a bond holding in the underlying company, it holds those bonds in both a bond- and equity-focused fund. The median holdings in bond-focused funds are about 8.5 times larger than those in the equity fund.

Bonds held in equity-focused funds drive the positive association between bond holdings and voting patterns. Table 5 shows this. Including both measures of bond holdings, we find that a one standard deviation increase in *Bond [in Equity Funds]/TNA* (0.003%) is associated with a 0.00507 percentage point increase in the likelihood of the institution voting against ISS, and the point estimate is statistically significant at the five percent level. Similar to Table 4, the coefficient remains about one-third of the observed coefficient for *Equity/TNA*. However, the coefficient on *Bond [in Bond Funds]/TNA* is 30% smaller and not statistically significant at conventional levels.

While these estimates suggest a difference in the relative importance of different bond holdings, we cannot reject the null hypothesis that the coefficients for *Bond [in Bond Funds]/TNA* and *Bond [in Equity Funds]/TNA* are the same. However, that changes when we repeat the estimation for different types of proposals, which we do in Table 6.

In Table 6, we repeat the estimation using all proposals and separately analyze the importance of bond holdings for contentious and non-contentious proposals and for contested and non-contested proposals. We define contested proposals as proposals where the vote outcome was within five percentage points of the threshold required for passage. If our findings reflect institutions being more attentive voters when a company's bonds represent a larger proportion of their portfolio, contentious and contested proposals, which are more likely to require investor attention, should drive these correlations rather than other proposals, which are more perfunctory.

We find that that association between bond holdings and voting is limited to contentious proposals and bond holdings found in equity-focused funds. When using all shareholder and manager proposals, both contentious and non-contentious, we find little association between voting patterns and holdings (Table 6, Column 1). The non-significant association is consistent with many of these proposals being routine management proposals that require little attention. However, when we restrict the sample to contentious proposals, we find that both bond and equity positions predict an increase in the likelihood of an institution disagreeing with ISS (Column 2). The association for bonds is limited to bond holdings in equity-focused funds; the coefficient on *Bond [in Bond Funds]/TNA* is negative and not statistically significant. There is no association between bond holdings and voting outcomes for non-contentious proposals (Column 3).[10]

We find a similar pattern when we divide the sample between contested and non-contested proposals. Bond holdings in equity-focused funds positively predict voting against ISS recommendations for contested proposals (Column 4). There is no association between holdings and institutions' voting when restricting the sample to non-contested proposals (Column 5).

Overall, these findings are consistent with bond holdings, particularly those in equity-focused funds, influencing institutions' level of attention to individual firms. Both larger equity and bond positions in a company increase an institution's likelihood of voting in ways that are positively associated with investor attention. These findings highlight that while only equity investors vote, an institution's holding of bonds might influence how attentive an equity owner is. This increased attention might occur for a variety of reasons. Fund managers who hold bond positions might possess additional information that influences an institution's decision on how to vote their equity positions. Moreover, because credit rating agencies increasingly factor in a company's activities on environmental, social, and governance (ESG) issues, those fund managers might encourage more active voting and monitoring. Because bond values are sensitive to long-

---

[10] The negative and statistically significant association between *Equity/TNA* and the likelihood of disagreeing with ISS for non-contentious proposals in Table 6, Column 3, indicates that institutions are also less likely to vote against ISS (and managers) when they hold a larger equity position and ISS agrees with the management recommendation (i.e., the proposal is non-contentious). We also find a negative coefficient for *Bond [in Equity Funds]/TNA* for such proposals, but the estimate is not statistically significant. However, both point estimates are an order of magnitude smaller than the coefficients observed for contentious proposals (Table 6, Column 2).

term tail risks, managers holding a bond position might also be more attentive to company-specific events critical to a company's viability and ability to meet long-term debt obligations.

4.3. Bond holdings and an institution's EDGAR viewings of company filings

Because voting against ISS need not always indicate an attentive investor, we also assess the association between bond holdings and an alternative proxy for investor attention—the number of times an institution accesses a company's proxy statement and other SEC filings via EDGAR in the days before a shareholder meeting. Because this proxy of investor attention is measured at the meeting- rather than proposal-level, we now estimate

$$Ln(1 + views)_{iklm} = \beta(\frac{Bond}{TNA})_{ikm} + \gamma(\frac{Equity}{TNA})_{ikm} + \alpha_l + \delta_{im} + \varepsilon_{iklm}, \quad (2)$$

where *views* is the number of times institution *i* accessed the EDGAR filings of firm *k* before shareholder meeting *l* held in month *m* (see Section 2.3 and the Appendix for more details on how we construct *views)*, and α_l and δ_{im} are meeting and institution-by-month fixed effects, respectively. We follow Iliev, Kalodimos, and Lowry (2020) and use the log of 1+*views* to mitigate the outcome variable's skewness while avoiding the loss of observations with zero views. We also report findings when using an indicator for non-zero *views* as the outcome of interest. We continue to cluster the standard errors at the fund family level.

Table 7 provides summary statistics for our meeting-by-institution-level sample. The average number of EDGAR views by an institution before a shareholder meeting in which the institution has a non-zero equity position is 0.112 views [i.e., $e^{0.106486} - 1$]. When the number of views is non-zero, the average is 2.64 views [i.e., $e^{1.29198} - 1$]. In 9.5% of observations, an institution also holds a bond position in the company at the time of the meeting. That bond position accounts for, on average, 0.0016% of the institution's overall portfolio. In 4% of observations, an institution holds a bond position in the company as part of an equity-focused fund, and in 8.3% of observations, they hold bonds in a bond-focused fund.[11]

[11] A combination of factors drives the lower proportion of observations with non-zero bond holdings in our meeting-by-institution sample (9.5%) relative to our proposal-by-institution sample (34.1%). First, the meeting-by-institution sample covers all meetings, not just those with contentious shareholder proposals. In a proposal-by-institution sample

Estimating eq. (2), we find that bond holdings predict how often an institution will access a company's proxy filing and other SEC filings via EDGAR in the days before the meeting. Table 8 reports our estimates. When excluding the control for *Equity/TNA*, an increase in *Bond/TNA* is associated with an increase in the number of EDGAR views (Column 1). The point estimate remains mostly unchanged when including *Equity/TNA* as a control (Column 2). Moreover, similar to Iliev, Kalodimos, and Lowry (2020), we find a positive association between institutions' equity holdings and the number of times institutions access a firm's SEC filings, consistent with institutions conducting more governance research on stocks that account for a larger proportion of their overall portfolio. Comparing the estimates for *Bond/TNA* and *Equity/TNA*, we see that a one standard deviation increase in *Bond/TNA* is associated with an increase in EDGAR views that is about 41.7% of the observed change in views for a one standard deviation increase in *Equity/TNA*.

A drawback of using EDGAR views as our outcome variable is that it does not allow us to focus on the shareholder proposals more likely to require investor attention. The sample used in Columns 1-2 of Table 8 includes many meetings with only routine proposals. To mitigate this weakness, we next assess whether the observed association varies when a meeting includes a contentious shareholder proposal. Table 8, Columns 3-4 conducts this test.

The association between bond holdings and EDGAR views is primarily driven by meetings that include a contentious shareholder proposal. When restricting the sample to meetings with a contentious shareholder proposal (which accounts for about 11% of all observations), we find a large, and positive coefficient on *Bond/TNA* (Column 3; $p < 0.05$). When using meetings without a contentious shareholder proposal, the point estimate is more than 50% smaller and no longer statistically significant (Column 4). The lack of an association between bond holdings and the number of EDGAR views for meetings without a contentious shareholder proposal is consistent with such meetings being more likely to include only routine proposals that require less attention.

---

that includes all proposals, the proportion of observations with non-zero bond holdings drops to 15.2%. This drop is because ISS is more likely to recommend voting against managers for larger companies, which are more likely to have publicly traded bonds. Second, the institutions that we can match IP addresses to are less likely to hold bonds relative to the institutions in the proposal-by-institution sample.

We find similar results when we instead analyze the likelihood of an institution accessing any filing before the meeting. A one standard deviation increase in *Bond/TNA* is associated with a 0.3 percentage increase in the likelihood of an institution accessing a filing before a meeting (Column 5). The finding is robust to controlling for *Equity/TNA* (Column 6). Similar to before, the association between bond holdings and EDGAR views is driven by meetings that include a contentious shareholder proposal (Column 7) rather than meetings without any such proposal (Column 8), and the *Bond/TNA* coefficient is 30% of the *Equity/TNA* coefficient's magnitude.

Bonds held in equity-focused funds continue to drive the association between bond holdings and investor attention. Table 9, where we repeat our estimation of eq. (2) but now separate the bond holdings into equity- and bond-focused funds, illustrates this pattern. Only the coefficient on bond holdings that are part of an equity fund is statistically significant. In terms of magnitude, bond holdings' relative importance is about one-third to one-half of *Equity/TNA*.

Overall, our findings are qualitatively similar when using an institution's EDGAR views before a meeting as an alternative proxy for investor attention. Using both proxies, we find that institutions are more likely to exhibit voting and online behaviors that indicate a more attentive voter when the underlying company's bonds account for a larger proportion of their portfolio, especially when those bonds are held in equity-focused funds.

5. Heterogeneity Across Institutions and Funds

We next assess whether the observed importance of bond holdings differs across indexed and actively managed funds. We also assess the importance of bond holdings for The Big Three institutions (i.e., BlackRock, State Street, and Vanguard), which account for about 75% of all indexed mutual fund and ETF assets in the US. For this analysis, we restrict our attention to our first proxy for attention, disagreeing with ISS, because it allows us to limit our sample to contentious shareholder proposals, where we observe more meaningful variation in attention.

5.1. Actively managed versus index funds

If bond holdings affect institutions' monitoring, particularly when these bonds are part of an equity-focused fund, the importance of these holdings might depend on the type of equity

fund—indexed or actively managed. If institutions are more attentive to their actively managed holdings, as found in Appel, Gormley, Keim, Kim, and Shin (2020), bonds held in indexed equity-focused funds (e.g., target-date funds) could matter less for institutions' monitoring than bonds held in actively managed equity-focused funds.

We further subdivide institutions' bond holdings held in equity-focused funds into bonds held in index funds and bonds held in actively managed funds to assess this possibility. To assign an equity fund as either indexed or actively managed, we follow Appel, Gormley, and Keim (2016, 2019) and classify a fund as "index" if either CRSP classifies the fund as indexed or if the fund name contains words such as "Index," "S&P," and "Russell," that would indicate an index fund. All other funds are classified as actively managed. About 21% of the funds in our sample are indexed, and actively managed funds hold about 93.2% of bond holdings found in equity funds.

Consistent with actively managed funds being more attentive to firm-specific shareholder proposals, we find that the positive association between bond holdings and the likelihood of voting against ISS is limited to bonds found in actively managed equity funds (Table 10). The amount of bonds held in index equity funds exhibits no association with whether an institution is likely to vote against ISS. The coefficient is negative and not statistically significant. The size of bond holdings held in non-indexed (i.e., actively managed) equity-focused funds is positively associated with the likelihood of disagreeing with ISS ($p < 0.01$). The relative importance of bond and equity holdings is similar to our earlier estimates. The point estimate for bond holdings in actively managed equity funds is about 40% of the point estimate obtained for *Equity/TNA*.

5.2. Bond holdings and the Big Three

We next assess whether bond holdings are associated with how actively the Big Three vote their shares. Because the index fund industry is highly concentrated, with the Big Three accounting for 75% of all indexed equity mutual fund and ETF assets, the growing popularity of indexing has resulted in the Big Three becoming some of the largest stock owners, and hence voters, in many US companies. In 2017, the Big Three cast around 25% of S&P 500 firms' votes, which account for about 75% of the total market capitalization for US public firms (Bebchuk and Hirst, 2019).

This growth of indexing and the importance of the Big Three has raised questions about how index investing affects corporate governance and whether the Big Three are motivated monitors. Although the monitoring of firms can help increase the value of these institutions' portfolios (Kahan and Rock 2019; Lewellen and Lewellen 2020), some argue that these institutions lack the incentive or firm-specific information required to monitor effectively (e.g., Schmidt and Fahlenbrach, 2018; Bebchuk and Hirst, 2019; Gilje, Gormley, and Levit, 2020). Despite this, evidence suggests these institutions exert influence over the companies they own (e.g., see Appel, Gormley, and Keim, 2016; Gormley, Gupta, Matsa, Mortal, and Yang, 2020).

The Big Three institutions, however, are also significant owners of bonds. Collectively, the Big Three account for about 40.1% of corporate bonds held in mutual funds and ETFs at the end of 2018, and corporate bonds accounted for, on average, about 13.5% of their total net assets. If bond holdings increase investor attention, then the Big Three's sizable holding of bonds might also contribute to their incentive to be engaged monitors. We assess this possibility in Table 11, where we repeat our estimation of eq. (1) and allow the importance of bonds to differ for the Big Three by including an interaction between *Bond/TNA* and an indicator that equals one if the voting institution is one of the Big Three institutions. Because of the additional interaction, the *Bond/TNA*'s coefficient captures the importance of bond holdings for all other institutions. In contrast, the sum of the coefficients on *Bond/TNA* and *Bond/TNA*× *Big 3 Indicator* captures the importance of bond holdings for the Big Three. For completeness, we also include an interaction term for *Equity/TNA* and the *Big 3 Indicator*. We do not include the *Big 3 Indicator* on its own, however, as it is collinear with our institution-by-month fixed effects.

The proportion of an institution's overall portfolio held in bonds is an even stronger predictor of voting against ISS for the Big Three institutions. For all other institutions, a one standard deviation increase in *Bond/TNA* is associated with a 0.00447 percentage point increase in the institution's likelihood of disagreeing with ISS (Table 11, Column 1; $p < 0.05$). For the Big Three, the same change in *Bond/TNA* is associated with an increase that is nearly five times larger (0.00447+0.0219 = 0.02637 percentage points) and significant at the one percent level.

A Big Three institution's likelihood of disagreeing with ISS is also more sensitive to *Equity/TNA* changes than other institutions. For all other institutions, a one standard deviation increase in *Equity/TNA* is associated with a 0.0139 percentage point increase in the institution's likelihood of disagreeing with ISS ($p < 0.01$). For the Big Three, the same change in *Equity/TNA* is associated with an increase that about 250% larger (0.0139+0.0366 = 0.0505 percentage points).

The greater importance of *Bond/TNA* and *Equity/TNA* for the Big Three's voting could reflect their portfolios' relative size and diversity. Because of their focus on indexed investment strategies, these institutions tend to hold more securities overall, and each security typically represents a relatively small proportion of their overall portfolio. Absent some economies of scale in monitoring, these institutions might focus their limited resources on monitoring companies that represent the largest proportion of their portfolio. Our findings in Table 11 are consistent with this possibility, which, to our knowledge, has not been shown before. Furthermore, we show that both the bond and equity positions of these institutions appear to matter.

We also find that even bonds held in index funds are associated with an increased likelihood of voting against ISS for the Big Three. Table 11, Column 2, where we separate bonds into those held in bond-focused funds, equity-focused index funds, and actively managed equity-focused funds, shows this finding. An increase in an institution's TNA held in bonds located in equity-focused index funds is positively associated with voting against ISS for the Big Three. The Big Three exhibit a similar-sized positive association between the proportion of bonds held in actively managed equity-focused funds and voting as other institutions. The proportion of net assets held in bond-focused funds continues to be unassociated with voting patterns.

Overall, these findings suggest that the Big Three's extensive bond holdings contribute to their incentive to monitor companies in their equity portfolio. Prior work has documented that large institutions, including the Big Three, have strong financial motives to monitor their equity investments because of the potential effect on fund fees and fund flows (Lewellen and Lewellen, 2020). However, this work ignores the possibility that active monitoring can also influence the value of their bond positions, providing them an additional motive to be engaged owners.

6. Robustness to Excluding Firms in Financial Distress

We next assess whether our findings are robust to excluding firms in financial distress, where debt and equity owners' interests might differ. Keswani, Tran, and Volpin (2020) find that institutions with dual debt and equity holdings are more likely to cast votes favorable to creditors, mainly when a firm is in financial distress. Because ISS recommendations reflect equity holders' interests, debt holding institutions' conflicting interests provide an alternative explanation for why such institutions are less likely to follow ISS recommendations.

However, our findings on voting are robust to excluding firms in financial distress. Table 12, where we repeat our earlier estimations but exclude firms that Keswani, Tran, and Volpin (2020) define as being distressed, shows this robustness.[12] Because the data needed to calculate financial distress is unavailable for some observations, we first repeat our baseline analysis on the subsample of observations with non-missing distress data. In this sample, which is about 90% of our original sample, we continue to find a positive association between institutions' bond holdings and the likelihood institutions vote against ISS (Column 1), especially for bond holdings in equity-focused funds (Column 3). Dropping firms currently in distress has minimal impact on the magnitudes (Columns 2 and 4). In unreported findings, we also find no evidence that the association between bond holdings and institutions' voting differs for firms in financial distress. These findings show that institutions' dual holdings matter more generally for how they vote and are not merely the result of creditor-shareholder conflicts.

Creditor-shareholder conflicts also cannot easily explain our findings for EDGAR viewings. While creditor-shareholder conflicts might induce a shift in voting, especially for firms in financial distress, it is unclear why it would explain the amount of governance research conducted by an institution. Consistent with this, our earlier findings for EDGAR viewings are nearly unchanged when excluding firms in financial distress (Appendix Table A1).

Overall, our findings suggest that institutions are more engaged monitors of their equity positions when they also hold a debt position. This finding provides an essential complement to

---

[12] Specifically, we exclude observations where the Bharath and Shumway (2008) distance to default measure for the firm indicates the firm's default probability is at least 75% in the year concerned.

the existing work on creditor-shareholder conflicts. While such conflicts might lead to votes that fail to maximize shareholder value when companies are in distress, more active monitoring can positively influence the value of both an institutions' debt and equity positions at other times.

7. Conclusion

Investors influence governance through a combination of voice (managerial engagement and voting; e.g., Shleifer and Vishny, 1986; Admati, Pfleiderer, and Zechner, 1994) and exit (selling one's position; e.g., Admati and Pfleiderer, 2009; Edmans, 2009). Lacking the ability to participate in shareholder votes, bond investors are typically not thought to play an important governance role. Nevertheless, bond investors have many reasons to be concerned about firms' governance structures, which can influence credit ratings and the likelihood of repayment. Moreover, bonds represent a large proportion of institutional investors' portfolios, providing bond investors a potential voice in how actively their institutions monitor and vote their shares.

We find evidence that institutions' bond holdings are associated with how attentive they are and the amount of governance research they conduct. Institutions are more likely to vote against ISS, an indication of active monitoring, and more likely access a company's SEC filings before a shareholder meeting, an indication of governance research, when they have a larger equity position in that company *and* when they have a larger bond position. Comparing the importance of the two holdings, an increase in the size of an institution's bond position is associated with increased active voting and governance research about one-third to one-half of the observed changes for a similarly sized increase in an institution's equity position.

Our findings highlight how the determinants of institutional investor attention can be complicated. Institutions do not just manage increasingly large equity portfolios; they also manage large bond portfolios. These combined holdings appear to play a factor in which institutions choose to allocate their limited attention and resources. Our findings also suggest that which type of funds hold these investments (e.g., equity- versus bond-focused) matters, as does the institution type. For example, bond positions are more associated with institutional attention when they are part of a fund that primarily focuses on equity.

Finally, our findings suggest that institutions' bond holdings increase their incentives to be engaged monitors, providing a potential counterpoint to recent concerns about how institutions' dual ownership might affect equity investors. While dual ownership of both a company's bonds and equity could increase the potential for voting decisions that benefit debt holders at the expense of equity investors (e.g., Bodnaruk and Rossi, 2016; Keswani, Tran, and Volpin, 2020), an overall increase in active monitoring and engagement could improve value for both investors. How these dual holdings and their increasing frequency among firms' largest institutional investors ultimately affect firms' governance structures is an important topic for future research.

Appendix A: Processing fund's information acquisition via EDGAR

The search traffic data for SEC.gov covers the period from February 2003 through June 2017. EDGAR log file data set includes information on visitor's Internet Protocol (IP) address, date, timestamp, CIK, and filing document's accession number. The IP address in the dataset is in version 4 (IPv4) format, which defines an IP address as a 32-bit number separated into four 8-bit numbers. A dot separates each 8-bit number, and the number between the dots could be between 0 and 255 (2^8 -1). So, a specific IP address, let us say BlackRock's, looks like 199.253.64.128. However, the last octet of the IP addresses in the EDGAR log files is replaced with alphabets. The replacement is done to preserve the uniqueness of the IP address and not reveal the visitor's full identity. Thus, if Blackrock accesses the SEC.gov website from the IP address, the log file will show an entry 199.253.64.gjs. In essence, the EDGAR log file dataset has a 24-bit (IP3) address for each EDGAR server activity. Fortunately, most fund families register large blocks of IP address; for example, BlackRock owns IP addresses ranging from 199.242.6.0 to 199.242.6.255. As such, the IP3 address is a sufficiently precise representative for IPv4 addresses.

Loughran, Mcdonalds (2017) suggests separating EDGAR requests generated by robots from server requests by regular investors. We classify an IP address as a robot if it requests more than a thousand filings in a day. We remove IP addresses classified as robots for that particular day. To include only valid EDGAR activities, we follow Drake, Roulstone, Thornock (2015), and exclude activities not related to governance research. We remove index pages (index.htm), icons (.ico), XML filings (.xml), and filings that are under 500 bytes in size. We also combine views by an IP address if they are less than five minutes apart and for the same filing.

The second part of our dataset is a lookup table from Digital Element, a geolocation data and services firm containing a timestamp of IP addresses (IPv4) and registered organization name as of December 2016. We use regular expressions, such as (.*blackrock.*) for BlackRock Financial Management, to get IPv4 associated with fund families. To assign IP3 blocks to fund families, we use a similar procedure as Iliev, Kalodimos, and Lowry (2020). If a fund family owns all or a subset of the IP3 address and no other fund family owns an address from the IP3 block, we attribute it to the fund family. If two or more fund families own a subset of IP3 block, we assign it to the family that contains the most IP address for the IP3 block. If two fund families own an equal number of IP addresses in an IP3 block, we drop those IP3 blocks. The chances of overestimating

views from assigning an entire IP3 block to a fund family if they own a fraction of addresses is low, as it is unlikely for non-financial firms to access filings from SEC.gov.

Next, we look for the validity of IP3 blocks assigned to the fund family. The IP address to the organization name lookup table is a snapshot from December 2016. However, fund families sometimes change their underlying technology infrastructure and, in that process, register for different IP3 blocks. To ensure that we have credible IP3 blocks, we go back quarterly from December 2016 and see what fraction of holdings fund family access through the EDGAR server. We use CRSP mutual fund data to get fund family holdings. If a fund family does not access more than 1% of its holding in two consecutive quarters, we stop including the fund family before the quarter. For example, Cambiar Investors accessed 1.9%, 3.3%, 0.0%, and 0.1% of its holdings in 2015Q4, 2015Q3, 2015Q2, and 2015Q1 respectively. Therefore, we exclude Cambiar Investors from our sample before June 2015.

Subsequently, we match valid IP3 blocks from the organization lookup table with IP3 from EDGAR log files. We identify proxy filings associated with a shareholder meeting (definitive proxy statement) based on the accession number of the filing in log files and SEC's index files. To measure the number of times a fund family accessed definitive proxy statements and other filings, we aggregate daily fund family views during the shareholders meeting window. We use a window starting from 30 days before the definitive proxy statement to the shareholder meeting date. The fund family's views, as measured from EDGAR log files, likely under-represents the actual views. As mentioned in Bauguess, Cooney, Hanley (2018), the EDGAR log files do not contain any SEC filings requests from EDGAR's FTP site. Moreover, internet service providers cache frequently requested documents for future ease of reference. So, requests for the same content that have been cached may not be captured by the log file.

References

Admati, Anat R. and Paul Pfleiderer, "The "Wall Street Walk" and Shareholder Activism: Exit as a Form of Voice," *The Review of Financial Studies*, July 2009, *22* (7), 2645–2685.

_ , _ , **and Josef Zechner**, "Large Shareholder Activism, Risk Sharing, and Financial Market Equilibrium," *Journal of Political Economy*, December 1994, *102* (6), 1097–1130.

Appel, Ian R., Todd A. Gormley, and Donald B. Keim, "Passive Investors, Not Passive Owners," *Journal of Financial Economics*, July 2016, *121* (1), 111–141.

_ , _ , **and** _ , "Standing on the Shoulders of Giants: The Effect of Passive Investors on Activism," *The Review of Financial Studies*, July 2019, *32* (7), 2720–2774.

Auh, Jun Kyung and Jennie Bai, "Cross-Asset Information Synergy in Mutual Fund Families," SSRN Scholarly Paper ID 3163135, Social Science Research Network, Rochester, NY November 2020.

Bauguess, Scott W., John Cooney, and Kathleen Weiss Hanley, "Investor Demand for Information in Newly Issued Securities," SSRN Scholarly Paper ID 2379056, Social Science Research Network, Rochester, NY June 2018.

Bebchuk, Lucian A and Scott Hirst, "The Specter of the Giant Three," Working Paper 25914, National Bureau of Economic Research June 2019.

Ben-Rephael, Azi, Zhi Da, and Ryan D. Israelsen, "It Depends on Where You Search: Institutional Investor Attention and Underreaction to News," *The Review of Financial Studies*, September 2017, *30* (9), 3009–3047.

Bharath, Sreedhar T. and Tyler Shumway, "Forecasting Default with the Merton Distance to Default Model," *The Review of Financial Studies*, May 2008, *21* (3), 1339–1369.

Bodnaruk, Andriy and Marco Rossi, "Dual Ownership, Returns, and Voting in Mergers," *Journal of Financial Economics*, April 2016, *120* (1), 58–80.

Brav, Alon, Wei Jiang, Tao Li, and James Pinnington, "Picking Friends Before Picking (Proxy) Fights: How Mutual Fund Voting Shapes Proxy Contests," SSRN Scholarly Paper ID 3101473, Social Science Research Network, Rochester, NY March 2019.

Chava, Sudheer, Rui Wang, and Hong Zou, "Covenants, Creditors' Simultaneous Equity Holdings, and Firm Investment Policies," *Journal of Financial and Quantitative Analysis*, April 2019, *54* (2), 481–512.

Chu, Yongqiang, "Shareholder-Creditor Conflict and Payout Policy: Evidence from Mergers between Lenders and Shareholders," *The Review of Financial Studies*, August 2018, *31* (8), 3098–3121.

_ **, Ha Nguyen, Jun Wang, Wei Wang, and Wenyu Wang**, "Simultaneous Debt-Equity Holdings and The Resolution of Financial Distress," *SSRN Electronic Journal*, 2020.

Dallas, George, "The Role of the Creditor in Corporate Governance and Investor Stewardship," October 2019.

Drake, Michael S., Darren T. Roulstone, and Jacob R. Thornock, "The Determinants and Consequences of Information Acquisition via EDGAR," *Contemporary Accounting Research*, 2015, *32* (3), 1128–1161.

Edmans, Alex, "Blockholder Trading, Market Efficiency, and Managerial Myopia," *The Journal of Finance*, 2009, *64* (6), 2481–2513.

Fang, Lily H., Joel Peress, and Lu Zheng, "Does Media Coverage of Stocks Affect Mutual Funds' Trading and Performance?," *The Review of Financial Studies*, December 2014, *27* (12), 3441–3466.

Fich, Eliezer M., Jarrad Harford, and Anh L. Tran, "Motivated Monitors: The Importance of Institutional Investors′ Portfolio Weights," *Journal of Financial Economics*, October 2015, *118* (1), 21–48.

Gantchev, Nickolay and Mariassunta Giannetti, "The Costs and Benefits of Shareholder Democracy: Gadflies and Low-Cost Activism," SSRN Scholarly Paper ID 3269378, Social Science Research Network, Rochester, NY September 2020.

Gilje, Erik P., Todd A. Gormley, and Doron Levit, "Who's Paying Attention? Measuring Common Ownership and Its Impact on Managerial Incentives," *Journal of Financial Economics*, July 2020, *137* (1), 152–178.

Gormley, Todd A., Zachary Kaplan, and Aadhaar Verma, "More Informative Disclosures, Less Informative Prices? Portfolio and Price Formation Around Quarter-Ends," SSRN Scholarly Paper ID 3067724, Social Science Research Network, Rochester, NY March 2020.

Gutiérrez, Germán and Thomas Philippon, "Ownership, Concentration, and Investment," *AEA Papers and Proceedings*, May 2018, *108*, 432–437.

Iliev, Peter and Michelle Lowry, "Are Mutual Funds Active Voters?," *The Review of Financial Studies*, February 2015, *28* (2), 446–485.

— , **Jonathan Kalodimos, and Michelle Lowry**, "Investors' Attention to Corporate Governance," SSRN Scholarly Paper ID 3162407, Social Science Research Network, Rochester, NY July 2020.

Jiang, Wei, Kai Li, and Pei Shao, "When Shareholders Are Creditors: Effects of the Simultaneous Holding of Equity and Debt by Non-Commercial Banking Institutions," *The Review of Financial Studies*, October 2010, *23* (10), 3595–3637.

Kahan, Marcel and Edward B. Rock, "Index Funds and Corporate Governance: Let Shareholders Be Shareholders," SSRN Scholarly Paper ID 3295098, Social Science Research Network, Rochester, NY April 2019.

Kedia, Simi, Laura T. Starks, and Xianjue Wang, "Institutional Investors and Hedge Fund Activism," SSRN Scholarly Paper ID 3560537, Social Science Research Network, Rochester, NY March 2020.

Kempf, Elisabeth, Alberto Manconi, and Oliver Spalt, "Distracted Shareholders and Corporate Actions," *The Review of Financial Studies*, May 2017, *30* (5), 1660–1695.

Keswani, Aneel, Anh Tran, and Paolo Volpin, "Institutional Debt Holder Governance," *Journal of Financial and Quantitative Analysis*, 2020, pp. 1–48.

Lewellen, Jonathan and Katharina Lewellen, "Institutional Investors and Corporate Governance: The Incentive to Be Engaged," SSRN Scholarly Paper ID 3265761, Social Science Research Network, Rochester, NY September 2020.

Lewellen, Katharina and Michelle Lowry, "Does Common Ownership Really Increase Firm Coordination?," SSRN Scholarly Paper ID 3336343, Social Science Research Network, Rochester, NY July 2020.

— **and** — , "Does Common Ownership Really Increase Firm Coordination?," SSRN Scholarly Paper ID 3336343, Social Science Research Network, Rochester, NY July 2020.

Liu, Claire, Angie Low, Ronald W. Masulis, and Le Zhang, "Monitoring the Monitor: Distracted Institutional Investors and Board Governance," *The Review of Financial Studies*, October 2020, *33* (10), 4489–4531.

Loughran, Tim and Bill McDonald, "The Use of EDGAR Filings by Investors," *Journal of Behavioral Finance*, April 2017, *18* (2), 231–248.

Lu, Yan, Sugata Ray, and Melvyn Teo, "Limited Attention, Marital Events and Hedge Funds," *Journal of Financial Economics*, December 2016, *122* (3), 607–624.

Malenko, Andrey and Nadya Malenko, "Proxy Advisory Firms: The Economics of Selling Information to Voters," *The Journal of Finance*, 2019, *74* (5), 2441–2490.

Potter, Mark E. and Christopher G. Schwarz, "Revisiting Mutual Fund Portfolio Disclosure," *The Review of Financial Studies*, December 2016, *29* (12), 3519–3544.

Schmidt, Cornelius and Rüdiger Fahlenbrach, "Do Exogenous Changes in Passive Institutional Ownership Affect Corporate Governance and Firm Value?," *Journal of Financial Economics*, May 2017, *124* (2), 285–306.

Schmidt, Daniel, "Distracted Institutional Investors," *Journal of Financial and Quantitative Analysis*, December 2019, *54* (6), 2453–2491.

Shleifer, Andrei and Robert W. Vishny, "Large Shareholders and Corporate Control," *Journal of Political Economy*, June 1986, *94* (3, Part 1), 461–488.

Yang, Huan, "Institutional Dual Holdings and Risk Shifting: Evidence from Corporate Innovation," SSRN Scholarly Paper ID 2837530, Social Science Research Network, Rochester, NY August 2019.

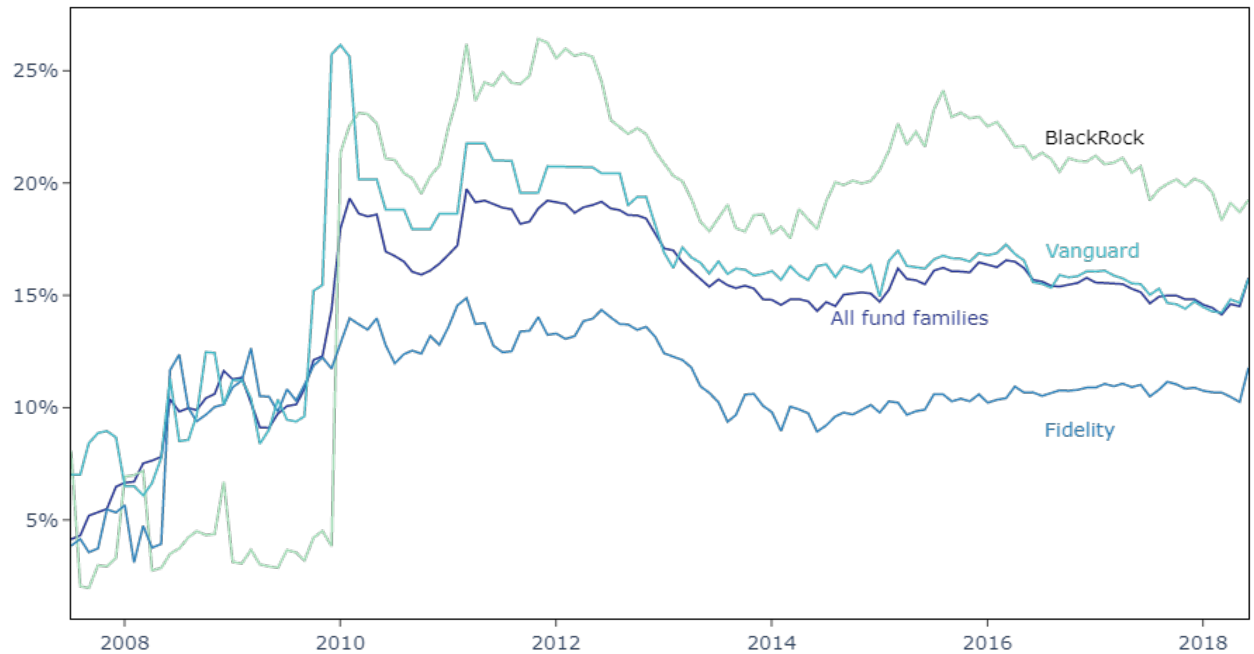


Figure 1 -- Institution's Bond Holdings as a Fraction of Total Net Assets, 2008-2018

This figure tabulates the weighted average share of total net assets held in bonds by quarter from 2008 to 2018 across all mutual fund families covered in the CRSP Mutual Fund Database. Separate plots are also provided for the three institutions with the largest total net assets at the end of 2018, Vanguard, Fidelity, and BlackRock.

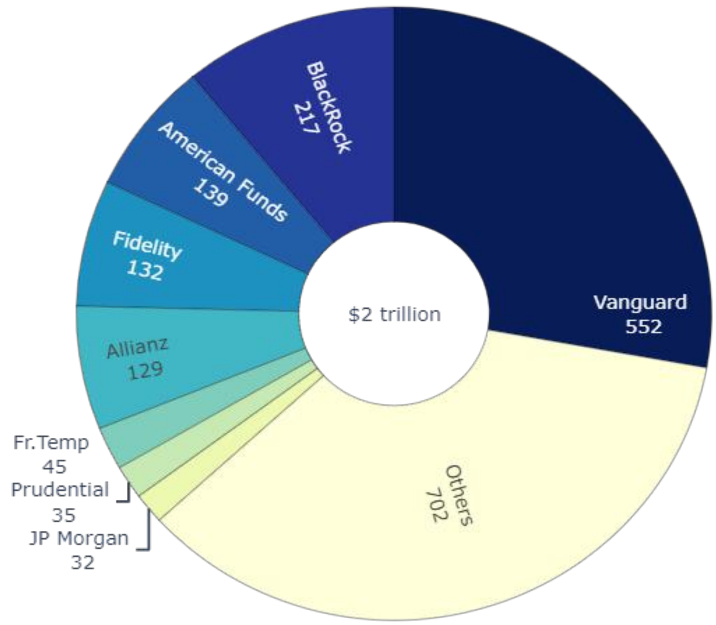


Figure 2 -- Corporate bond holdings by fund families as of December 2018

This figure plots the corporate bond holdings of the top eight fund families at the end of 2018. The number next to the fund family indicates corporate bond holdings in USD billion. Total corporate bond holding by mutual fund institutions is annotated in the center.

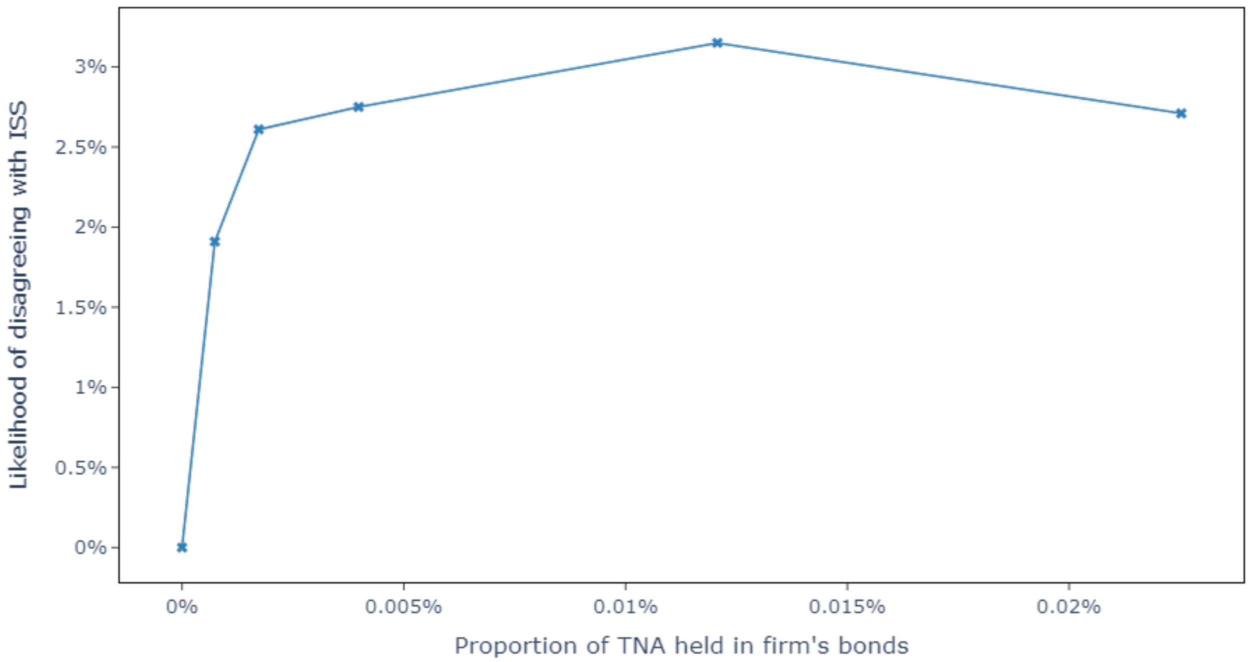


Figure 3 -- Non-Parametric Estimation of Association Between Voting Against ISS and Bond/TNA

This figure plots the point estimates from the proposal-by-institution-level regression of an indicator for disagreement with ISS onto quintile indicatrs for each *Bond/TNA* quintile. The regression includes proposal and institution-by-month fixed effects, and a linear extrapolation is applied between point estimates to construct the figure, where the likelihood of disagreeing with ISS is centered at zero for *Bond/TNA* = 0.

Table 1 -- Bond & Equity Holdings of Fund Families

This table tabulates the total net assets (in $ billions) and breakdown of these assets between bonds and equity for all mutual fund families, the six largest non-index fund families, and the "Big 3" index fund families as of December 2018, as calculated using the CRSP Mutual Fund data.

	Total net assets (TNA) in $ billions	Equity % of TNA	Bond % of TNA
All Mutual Fund Families	12,571	84%	16%
Six Largest, non-Index Fund Families			
Fidelity	1,119	88%	12%
American Funds	1,066	87%	13%
T Rowe Price	397	92%	8%
Dimensional Fund Advisors	253	89%	11%
Invesco	249	91%	9%
Allianz	225	42%	58%
The "Big Three" Index Fund Families			
Vanguard	3,503	84%	16%
BlackRock	1,128	81%	19%
State Street	499	95%	5%

Table 2 -- Mutual fund holdings by year and Lipper classifications

This table provides a breakdown of mutual fund holdings and Lipper asset classifications by year from 2008 to 2018. Columns (1)-(3) tabulate the percent of funds with equity, bond, and mixed holdings by year. Column (4) tabulates the average percent of total net assets (TNA) held in equity for funds with both bond and equity positions, while column (5) tabulates the percent of these funds that are classified as "equity" funds by Lipper, as provided in CRSP. For funds classified as "equity" funds, column (6) tabulates the percent with bond positions, and column (7) tabulates the average proportion of TNA that is held in bonds for "equity" funds with non-zero bond holdings. Columns (8)-(9) tabulate similar statistics for funds classified as "debt" funds by Lipper.

	% of funds with			For funds with both bonds and equity		For funds classified as "equity" by Lipper		For funds classified as "debt" by Lipper	
Year	Only equity	Only bonds	Both	Avg. % of TNA in equity	% classified as “equity” funds by Lipper	% with bonds in them	Avg. bond % of TNA when hold bonds	% with equity in them	Avg. equity % of TNA when hold equity
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2008	83%	2%	16%	47%	44%	8%	20%	87%	25%
2009	78%	2%	20%	49%	48%	11%	16%	86%	17%
2010	71%	6%	23%	56%	61%	17%	16%	62%	13%
2011	66%	10%	25%	60%	66%	21%	18%	51%	12%
2012	66%	11%	24%	60%	67%	21%	20%	46%	12%
2013	65%	11%	24%	60%	67%	21%	20%	46%	14%
2014	66%	11%	24%	60%	66%	20%	19%	47%	17%
2015	65%	11%	24%	58%	65%	20%	20%	48%	18%
2016	66%	11%	23%	56%	64%	19%	24%	47%	19%
2017	70%	11%	19%	48%	56%	14%	30%	48%	19%
2018	70%	11%	19%	46%	53%	14%	31%	51%	21%
Average	69%	9%	22%	55%	61%	17%	21%	52%	17%

Table 3 -- Summary Statistics for Proposal-by-Institution Sample

This table presents summary statistics for our proposal-by-institution-level outcome and explanatory variables. To match our later estimations, the sample is limited to contentious shareholder proposals (i.e., proposals where ISS recommended voting against management) that were voted on between 2008 and 2018. The likelihood of voting against ISS is an indicator for institution *i* voting against the ISS recommendation on proposal *j* for firm *k* in month *m*, *Bond Holdings / TNA* is the proportion of institution *i*'s total net assets (TNA) that is held in firm *k*'s bonds as of month *m, and Equity Holdings / TNA is the* fraction of the institution's TNA that is held in firm *k*'s stock. In cases where an institution's funds do not all cast the same type of vote, we define the voting against ISS variable as the proportion of voting funds that vote against the ISS recommendation. *Bond Holdings [in Bond Funds] / TNA* and *Bond Holdings [in Equity Funds] / TNA* are the proportion of institution *i*'s total net assets (TNA) that is held in firm *k*'s bonds via funds classified as bond and equity funds by CRSP.

	Mean	Median	SD	% of observations with non-zero value	Mean if non-zero	Number
Likelihood of Voting Against ISS	0.40242800	0	0.46104150	47.3%	0.84995910	276,566
Bond Holdings / TNA	0.00011850	0	0.00030150	34.1%	0.00034750	276,566
Equity Holdings / TNA	0.00252690	0.0006750	0.00434810	100%	0.00252690	276,566
Bond Holdings [in Bond Funds] / TNA	0.00009930	0	0.00025870	31.4%	0.00031610	276,566
Bond Holdings [in Equity Funds] / TNA	0.00000803	0	0.00002760	16.0%	0.00005010	276,566

Table 4 -- Voting Against ISS and Bond Holdings

This table presents coefficients from a proposal-by-institution-level estimation that regresses the likelihood of an institutional investor voting against the ISS recommendation for a given proposal onto measures of how important that proposal's company's bonds are in the overall portfolio of the institution. Specifically, we estimate the following:

$$Against_{ijkm} = \beta(\frac{Bond}{TNA})_{ikm} + \alpha_j + \delta_{im} + \varepsilon_{ijkm},$$

where *Against* is an indicator for a institution *i* voting against the ISS recommendation on proposal *j* for firm *k* in month *m*, *Bond/TNA* is the proportion of institution *i*'s total net assets (TNA) that is held in firm *k*'s bonds as of month *m* scaled by its sample standard deviation, and α_j and δ_{im} are proposal and institution-by-month fixed effects, respectively. In cases where an institution's funds do not all cast the same type of vote, we define *Against* as the proportion of voting funds that vote against the ISS recommendation. In column (2), the estimation also includes a control for the fraction of the institution's TNA that is held in firm *k*'s stock scaled by its sample standard deviation. The sample is limited to contentious shareholder proposals (i.e., proposals where ISS recommended voting against management) that were voted on between 2008 and 2018. *t*-statistics are reported in parentheses, and the standard errors are clustered by fund family, and ***, **, and * reflect statistical significance at the 1, 5, and 10% confidence intervals, respectively.

	Likelihood of Voting Against ISS	
	(1)	(2)
Bond Holdings / TNA	0.00546**	0.00524**
	(2.42)	(2.33)
Equity Holdings / TNA		0.0153***
		(4.12)
Institution-by-Month Fixed Effects	X	X
Proposal Fixed Effects	X	X
N	274,266	274,266
R-sq	0.555	0.555

Table 5 -- Voting Against ISS and Bond Holdings in "Bond" versus "Equity" Funds

This table presents coefficients from a proposal-by-institution-level estimation that regresses the percent of an institution's funds that vote against the ISS recommendation for a given proposal onto measures of how important that proposal company's bonds and equity are in the overall portfolio of the institution. Specifically, we estimate:

$$Against_{ijkm} = \beta_1\left(\frac{Bond\ [in\ Bond\ Funds]}{TNA}\right)_{ikm} + \beta_2\left(\frac{Bond\ [in\ Equity\ Funds]}{TNA}\right)_{ikm} + \theta\left(\frac{Equity}{TNA}\right)_{ikm} + \alpha_j + \delta_{im} + \varepsilon_{ijkm},$$

where *Against* is an indicator for a institution i voting against the ISS recommendation on proposal j for firm k in month m, *Bond [in Bond Funds]/TNA* and *Bond [in Equity Funds]/TNA* are the proportion of institution i's total net assets (TNA) that is held in firm k's bonds via funds classified as bond and equity funds by CRSP as of month m scaled by their sample standard deviations, *Equity/TNA* is the proprotion of the institution's TNA held in firm k's stock scaled by its sample standard deviation, and α_j and δ_{im} are proposal and institution-by-month fixed effects, respectively. In cases where an institution's funds do not all cast the same type of vote, we define *Against* as the proportion of funds that vote against the ISS recommendation. The sample is limited to contentious shareholder proposals (i.e., proposals where ISS recommended voting against management) that were voted on between 2008 and 2018. t-statistics are reported in parentheses, and the standard errors are clustered by institution, and ***, **, and * reflect statistical significance at the 1, 5, and 10% confidence intervals, respectively.

	% Funds Voting Against ISS
	(1)
Bond Holdings [in Bond Funds] / TNA	0.00358
	(1.41)
Bond Holdings [in Equity Funds] / TNA	0.00507**
	(2.17)
Equity Holdings / TNA	0.0152***
	(4.10)
Institution-by-Month Fixed Effects	X
Proposal Fixed Effects	X
N	274,266
R-sq	0.555

Table 6 -- Bond Holdings, Voting Against ISS, Type of Proposal

This table presents coefficients from a proposal-by-institution-level estimation that regresses the percent of an institution's funds that vote against the ISS recommendation for a given proposal onto measures of how important that proposal's company's bonds and equity are in the overall portfolio of the institution. Specifically, we estimate the following:

$$Against_{ijkm} = \beta_1\left(\frac{Bond\ [in\ Bond\ Funds]}{TNA}\right)_{ikm} + \beta_2\left(\frac{Bond\ [in\ Equity\ Funds]}{TNA}\right)_{ikm} + \theta\left(\frac{Equity}{TNA}\right)_{ikm} + \alpha_j + \delta_{im} + \varepsilon_{ijkm},$$

where *Against* is an indicator for a institution i voting against the ISS recommendation on proposal j for firm k in month m, *Bond [in Bond Funds]/TNA* and *Bond [in Equity Funds]/TNA* are the proportion of institution i's total net assets (TNA) that is held in firm k's bonds via funds classified as bond and equity funds by CRSP as of month m scaled by their sample standard deviation, *Equity/TNA* is the proprotion of the institution's TNA held in firm k's stock scaled by its sample standard deviation, and α_j and δ_{im} are proposal and institution-by-month fixed effects, respectively. In cases where an institution's funds do not all cast the same type of vote, we define *Against* as the proportion of funds that vote against the ISS recommendation. In column (1), the estimation includes all proposals rather than just contentious shareholder proposals as in earlier tables. Column (2) restricts the sample to contentious management and shareholder proposals (i.e., proposals where ISS reccomends voting against management), while column (3) restricts the sample to non-contentious proposals. Column (4) restricts the sample to contested proposals, which are defined as proposals where the final vote outcomes was within five percentage points of passage/failure, and column (5) restricts the sample to non-contested proposals. The sample is limited to proposals that were voted on between 2008 and 2018. *t*-statistics are reported in parentheses, and the standard errors are clustered by institution, and ***, **, and * reflect statistical significance at the 1, 5, and 10% confidence intervals, respectively.

	% Funds Voting Against ISS				
	All Proposals	Contentious	Not Contentious	Contested	Uncontested
	(1)	(2)	(3)	(4)	(5)
Bond Holdings [in Bond Funds] / TNA	-0.000212	-0.000545	-0.000307	-0.00161	-0.000178
	(-0.21)	(-0.36)	(-0.29)	(-1.45)	(-0.18)
Bond Holdings [in Equity Funds] / TNA	0.000461	0.00710***	-0.000418	0.00349***	0.000380
	(0.88)	(3.30)	(-0.81)	(2.68)	(0.74)
Equity Holdings / TNA	0.000472	0.0151***	-0.00125**	0.00415*	0.000367
	(0.78)	(5.43)	(-2.06)	(1.82)	(0.61)
Institution-by-Month Fixed Effects	X	X	X	X	X
Proposal Fixed Effects	X	X	X	X	X
N	13,395,141	1,346,932	12,046,562	116,868	13,275,284
R-sq	0.361	0.468	0.319	0.437	0.363

Table 7 -- Summary Statistics for Meeting-by-Institution Sample

This table presents summary statistics for our meeting-by-institution-level outcome and explanatory variables. The sample is limited to meetings that occurred from 2008 to 2018 and the mutual fund families that had a non-zero equity position in the company in the month of that meeting. *Ln(1+views)* is the natural log of one plus the number of times fund family *f* viewed the proxy filings for firm *k* prior to meeting *l* held by that firm in month *m*, *Bond/TNA* and *Equity/TNA* are the proportion of institution *i*'s total net assets (TNA) that is held in firm *k*'s bonds and equity as of month *m*. *Bond Holdings [in Bond Funds] / TNA* and *Bond Holdings [in Equity Funds] / TNA* are the proportion of institution *i*'s total net assets (TNA) that is held in firm *k*'s bonds via funds classified as bond and equity funds by CRSP.

	Mean	Median	SD	% of observations with non-zero value	Mean if non-zero	Number
Log(1 + EDGAR views of proxy filings)	0.10648600	0	0.41115720	8.2%	1.29198000	1,243,680
Bond Holdings / TNA	0.00001480	0	0.00007370	9.5%	0.00015620	1,243,680
Equity Holdings / TNA	0.00060050	0.0000639	0.00159480	100.0%	0.00060050	1,243,680
Bond Holdings [in Bond Funds] / TNA	0.00001220	0	0.00006300	8.3%	0.00014640	1,243,680
Bond Holdings [in Equity Funds] / TNA	0.00000106	0	0.00000712	4.0%	0.00002630	1,243,680

Table 8 -- Bond Holdings and Institutions' Viewing of Company Filings Prior to Meetings

This table presents coefficients from a meeting--by-institution-level estimation that regresses the number of times the fund family viewed a company's filings via EDGAR prior to a vote onto measures of how important that company's bonds and equity are in the overall portfolio of the fund family. Specifically, we estimate the following:

$$Ln(1+views)_{iklm} = \beta(\frac{Bond}{TNA})_{ikm} + \theta(\frac{Equity}{TNA})_{ikm} + \gamma_l + \delta_{im} + \varepsilon_{iklm},$$

where *Ln(1+views)* is the natural log of one plus the number of times institution *i* viewed the proxy filings for firm *k* prior to meeting *l* held by that firm in month *m*, *Bond/TNA* and *Equity/TNA* are the proportion of institution *i*'s total net assets (TNA) that is held in firm *k*'s bonds and equity as of month *m* scaled by their sample standard deviation, and γ_l and δ_{im} are meeting and institution-by-month fixed effects, respectively. We follow Iliev, Kalodimos, and Lowry (2018) in identifying and counting the number of times a fund family accessed a firm's filings via EGDAR; details are provided in the appendix. Columns (1)-(4) use *Ln(1+views)* as the dependent variable, while columns (5)-(8) use an indicator for non-zero views as the dependent variable. In columns (1)-(2) and (5)-(6), the sample includes all meetings that were held between 2008 and 2018 where the institution held some equity in the company prior to the meeting. Columns (3) and (7) restrict the sample to meetings with at least one contentious shareholder proposal, while columns (4) and (8) restrict the sample to meetings with no contentious shareholder proposal. *t*-statistics are reported in parentheses, and the standard errors are clustered by institution, and ***, **, and * reflect statistical significance at the 1, 5, and 10% confidence intervals, respectively.

	Log(1 + EDGAR views of filings)				Indicator for non-zero EDGAR views			
	All Meetings		Meetings with contentious shareholder proposal	Meetings with no contentious shareholder proposals	All Meetings		Meetings with contentious shareholder proposal	Meetings with no contentious shareholder proposals
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Bond Holdings / TNA	0.00579**	0.00551**	0.00600**	0.00288	0.00296**	0.00277**	0.00267**	0.00150
	(2.10)	(2.02)	(2.28)	(1.46)	(2.30)	(2.19)	(2.28)	(1.54)
Equity Holdings / TNA		0.0132***	0.0147***	0.0109***		0.00868***	0.00938***	0.00745***
		(4.07)	(4.09)	(3.68)		(4.31)	(5.01)	(3.57)
Institution-by-Month Fixed Effects	X	X	X	X	X	X	X	X
Meeting ID Fixed Effects	X	X	X	X	X	X	X	X
N	1,241,469	1,241,469	137,520	1,103,946	1,241,469	1,241,469	137,520	1,103,946
R-sq	0.223	0.224	0.329	0.208	0.242	0.242	0.322	0.231

Table 9 --EDGAR Viewings and "Bond" versus "Equity" Funds

This table presents coefficients from a meeting--by-institution-level estimation that regresses the number of times the institution viewed a company's proxy documents via EDGAR prior to a vote onto measures of how important that company is in the overall bond and equity portfolio of the institution. Specifically, we estimate the following:

$$Ln(1+views)_{iklm} = \beta_1(\frac{Bond\ (in\ Bond\ Funds)}{TNA})_{ikm} + \beta_2(\frac{Bond\ (in\ Equity\ Funds)}{TNA})_{ikm} + \theta(\frac{Equity}{TNA})_{ikm} + \gamma_l + \delta_{im} + \varepsilon_{iklm},$$

where *Ln(1+views)* is the natural log of one plus the number of times institution *i* viewed the proxy filings for firm *k* prior to meeting *l* held by that firm in month *m*, , *Bond [in Bond Funds]/TNA* and *Bond [in Equity Funds]/TNA* are the proportion of institution *i*'s total net assets (TNA) that is held in firm *k*'s bonds via funds classified as bond and equity funds by CRSP as of month *m* scaled by their sample standard deviation, *Equity/TNA* is the proprotion of the institution's TNA held in firm *k*'s stock scaled by its sample standard deviation, and γ_l and δ_{im} are meeting and institution-by-month fixed effects, respectively. We follow Iliev, Kalodimos, and Lowry (2018) in identifying and counting the number of times an institution accessed a firm's proxy filings via EGDAR; details are provided in the appendix. Column (1) use *Ln(1+views)* as the dependent variable, while column (2) uses an indicator for non-zero views as the dependent variable. The sample is limited to meetings that were held between 2008 and 2018 where the institution held some equity in the company at the time of the meeting. *t*-statistics are reported in parentheses, and the standard errors are clustered by institution, and ***, **, and * reflect statistical significance at the 1, 5, and 10% confidence intervals, respectively.

	Log(1 + EDGAR views of filings)	Indicator for non-zero EDGAR views
	(1)	(2)
Bond Holdings [in Bond Funds] / TNA	0.00341	0.00169
	(1.14)	(1.23)
Bond Holdings [in Equity Funds] / TNA	0.00628**	0.00316**
	(2.06)	(2.11)
Equity Holdings / TNA	0.0131***	0.00866***
	(4.08)	(4.31)
Institution-by-Month Fixed Effects	X	X
Meeting ID Fixed Effects	X	X
N	1,241,469	1,241,469
R-sq	0.224	0.242

Table 10 -- Voting, Bond Holdings, and Indexed versus Actively Managed Funds

This table presents coefficients from a proposal-by-institution-level estimation that regresses the percent of an institution's funds that vote against ISS for a given proposal onto measures of how important that proposal's company is in the overall portfolio of the institution. Specifically, we estimate:

$$Against_{ijkm} = \beta_1(\frac{Bond\ [in\ Bond\ Funds]}{TNA})_{ikm} + \beta_2(\frac{Bond\ [in\ non\text{-}Index\ Equity\ Funds]}{TNA})_{ikm} + \beta_3(\frac{Bond\ [in\ Index\ Equity\ Funds]}{TNA})_{ikm} + \theta(\frac{Equity}{TNA})_{ikm} + \alpha_j + \delta_{im} + \varepsilon_{ijkm},$$

where *Against* is an indicator for a institution i voting against the ISS recommendation on proposal j for firm k in month m, *Bond [in Bond Funds]/TNA* is the proportion of institution i's total net assets (TNA) that is held in firm k's bonds via funds classified as bond funds by CRSP as of month m, *Bond [in non-Index Funds]/TNA* and *Bond [in Index Equity Funds]/TNA* are the proportion of institution i's total net assets (TNA) that are held in firm k's bonds via funds classified as equity funds by CRSP broken down by whether the fund is an index fund or not, where each variable is scaled by its sample standard deviation, *Equity/TNA* is the proprotion of the institution's TNA held in firm k's stock scaled by its sample standard deviation, and α_j and δ_{im} are proposal and fund-family-by-month fixed effects, respectively. In cases where an institution's funds do not all cast the same type of vote, we define *Against* as the proportion of funds that vote against the ISS recommendation. The sample is limited to contentious shareholder proposals (i.e., proposals where ISS recommended voting against management) that were voted on between 2008 and 2018. t-statistics are reported in parentheses, and the standard errors are clustered by fund family, and ***, **, and * reflect statistical significance at the 1, 5, and 10% confidence intervals, respectively.

	% Funds Voting Against ISS
	(1)
Bond Holdings [in Bond Funds] / TNA	0.00332
	(1.30)
Bond Holdings [in non-Index Equity Funds] / TNA	0.00601***
	(2.81)
Bond Holdings [in Index Equity Funds] / TNA	-0.000937
	(-0.34)
Equity Holdings / TNA	0.0151***
	(4.08)
Institution-by-Month Fixed Effects	X
Proposal Fixed Effects	X
N	274,266
R-sq	0.555

Table 11 -- Bond Holdings, Voting Against ISS, and the "Big Three"

This table presents coefficients from a proposal-level estimation that regresses the percent of an institution's funds that vote against the ISS recommendation for a given proposal onto measures of how important that proposal's company is in the overall bond and equity portfolio of the fund family but now allows the association to vary for the "Big Three" institutions (Vanguard, BlackRock, State Street). Specifically, the estimations is the same as Column (1) of Tables 5 and 10, except that we now create an indicator that flags the three largest index providers, Big Three, and interact it with each of the explanatory variables. *Against* is an indicator for a institution i voting against the ISS recommendation on proposal j for firm k in month m, *Bond [in Bond Funds]/TNA* is the proportion of institution i's total net assets (TNA) that is held in firm k's bonds via funds classified as bond funds by CRSP as of month m scaled by its sample standard deviation, *Bond [in non-Index Funds]/TNA* and *Bond [in Index Equity Funds]/TNA* are the proportion of instituion i's total net assets (TNA) that are held in firm k's bonds via funds classified as equity funds by CRSP broken down by whether the fund is an index fund or not, where each is scaled by its sample standard deviation, *Equity/TNA* is the proprotion of fund family's TNA held in firm k's stock scaled by its sample standard deviation, and α_j and δ_{im} are proposal and institution-by-month fixed effects, respectively. In cases where an institution's funds do not all cast the same type of vote, we define *Against* as the proportion of funds that vote against the ISS recommendation. The sample is limited to contentious shareholder proposals (i.e., proposals where ISS recommended voting against management) that were voted on between 2008 and 2018. t-statistics are reported in parentheses, and the standard errors are clustered by fund family, and ***, **, and * reflect statistical significance at the 1, 5, and 10% confidence intervals, respectively.

	% Funds Voting Against ISS	
	(1)	(2)
Bond Holdings / TNA	0.00447**	
	(2.05)	
Bond Holdings / TNA × "Big Three" Indicator	0.0219***	
	(6.16)	
Bond Holdings [in Bond Funds] / TNA		0.00304
		(1.19)
Bond Holdings [in Bond Funds] / TNA × "Big Three" Indicator		0.00769
		(0.60)
Bond Holdings [in non-Index Equity Funds] / TNA		0.00522**
		(2.55)
Bond Holdings [in non-Index Equity Funds] / TNA × "Big Three" Indicator		0.00474
		(1.04)
Bond Holdings [in Index Equity Funds] / TNA		-0.00331*
		(-1.74)
Bond Holdings [in Index Equity Funds] / TNA × "Big Three" Indicator		0.0126***
		(3.85)
Equity Holdings / TNA	0.0139***	0.0138***
	(3.70)	(3.67)
Equity Holdings / TNA × "Big Three" Indicator	0.0366***	0.0352***
	(6.23)	(6.26)
Institution-by-Month Fixed Effects	X	X
Proposal Fixed Effects	X	X
N	274,266	274,266
R-sq	0.555	0.556

Table 12 -- Robustness to Excluding Firms in Financial Distress

This table presents coefficients from a proposal-by-institution-level estimation that regresses the likelihood of an institutional investor voting against the ISS recommendation for a given proposal onto measures of how important that proposal's company's bonds are in the overall portfolio of the institution. The estimation is the same as in Table 4 (Column 2) and Table 5, except that the sample is restricted to observations with the data necessary to calculate a firm's distance to default at the time of the vote, where distance to default is calculated using the approach of Bharath and Shumway (2008). Columns 2 and 4 further exclude firms where this distance to default measure indicates a firm's default probability is at least 75 percent, which is the threshold used in Keswani, Tran, and Volpin (2020) to flag financially distressed firms. *t*-statistics are reported in parentheses, and the standard errors are clustered by fund family, and ***, **, and * reflect statistical significance at the 1, 5, and 10% confidence intervals, respectively.

	Likelihood of Voting Against ISS			
	(1)	(2)	(3)	(4)
Bond Holdings / TNA	0.00533**	0.00477**		
	(2.37)	(2.24)		
Bond Holdings [in Bond Funds] / TNA			0.00384	0.00327
			(1.49)	(1.36)
Bond Holdings [in Equity Funds] / TNA			0.00460*	0.00474**
			(1.95)	(2.14)
Equity Holdings / TNA	0.0157***	0.0159***	0.0156***	0.0158***
	(4.26)	(4.31)	(4.24)	(4.29)
Institution-by-Month Fixed Effects	X	X	X	X
Proposal Fixed Effects	X	X	X	X
Sample Excludes Firms in Financial Distress		X		X
Sample Restricted to Obs. w/ Non-missing Distress Data	X	X	X	X
N	245,832	239,516	245,832	239,516
R-sq	0.557	0.559	0.558	0.559

Appendix

Table A1 -- Robustness of EDGAR Findings to Excluding Firms in Financial Distress

This table presents coefficients from a meeting-by-institution-level estimation that regresses the number of times the institution viewed a company's proxy documents via EDGAR prior to a vote onto measures of how important that company is in the overall bond and equity portfolio of the institution. The estimation and sample is the same as in Table 8 (Columns 2 and 6) and Table 9, except that the sample is restricted to observations with the data necessary to calculate a firm's distance to default at the time of the meeting, where distance to default is calculated using the approach of Bharath and Shumway (2008). Columns 2, 4, 6, and 8 further exclude firms where this distance to default measure indicates a firm's default probability is at least 75 percent, which is the threshold used in Keswani, Tran, and Volpin (2020) to flag financially distressed firms. *t*-statistics are reported in parentheses, and the standard errors are clustered by fund family, and ***, **, and * reflect statistical significance at the 1, 5, and 10% confidence intervals, respectively.

	Log(1 + EDGAR views of filings)				Indicator for non-zero EDGAR views			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Bond Holdings / TNA	0.00520*	0.00503*			0.00247*	0.00247*		
	(1.93)	(1.95)			(1.92)	(1.96)		
Bond Holdings [in Bond Funds] / TNA			0.00340	0.00325			0.00163	0.00162
			(1.17)	(1.17)			(1.19)	(1.21)
Bond Holdings [in Equity Funds] / TNA			0.00586**	0.00570**			0.00273**	0.00266**
			(2.01)	(2.01)			(2.00)	(1.98)
Equity Holdings / TNA	0.0133***	0.0133***	0.0132***	0.0132***	0.00891***	0.00893***	0.00888***	0.00890***
	(3.96)	(3.98)	(3.96)	(3.98)	(4.23)	(4.23)	(4.22)	(4.23)
Institution-by-Month Fixed Effects	X	X	X	X	X	X	X	X
Meeting ID Fixed Effects	X	X	X	X	X	X	X	X
Sample Excludes Firms in Financial Distress		X		X		X		X
Sample Restricted to Obs. w/ Non-missing Distress Data	X	X	X	X	X	X	X	X
N	965,176	945,670	965,176	945,670	965,176	945,670	965,176	945,670
R-sq	0.300	0.301	0.300	0.301	0.342	0.343	0.342	0.343

Does finance benefit society?

A language embedding approach

Manish Jha, Hongyi Liu, and Asaf Manela

Washington University in St. Louis

October 2020

Abstract

We measure popular sentiment toward finance using a computational linguistics approach applied to millions of books published in eight countries over hundreds of years. We document persistent differences in finance sentiment across countries despite ample time-series variation. Finance sentiment declines after epidemics and earthquakes, but rises following droughts, floods, and landslides. These heterogeneous effects of natural disasters suggest finance sentiment responds differently to the realization of insured versus uninsured risks. Using local projections, we find that positive shocks to finance sentiment have positive and persistent effects on economic growth. Our estimates predict a contraction in finance sentiment due to the COVID-19 pandemic that will exacerbate its long-term economic damage.

Keywords: sentiment, natural disasters, text analysis, word embedding, COVID-19

We thank Yves-Paul Auffray, Oscar Iglesias, Jay Kloecker, David Lindequist, Rodrigo Moser, Andrea Paloschi, Rita Podledneva, Tatiana Vdovina, Juan Ignacio Vizcaino, and seminar participants at the Virtual Finance Seminar and Washington University in St. Louis for helpful comments. E-mail: mjha@wustl.edu, hongyi.liu@wustl.edu, or amanela@wustl.edu. Computations used the Washington University Center for High Performance Computing, which is partially funded by NIH grant S10 OD018091.

``As finance academics, we should care deeply about the way the financial industry is perceived by society. Not so much because this affects our own reputation, but because there might be some truth in all these criticisms, truths we cannot see because we are too embedded in our own world. And even if we thought there were no truth, we should care about the effects that this reputation has in shaping regulation and government intervention in the financial industry. Last but not least, we should care because the positive role that finance can play in society depends on the public's perception of our industry.''

(Zingales, 2015, AFA Presidential Address)

1 Introduction

Positive popular sentiment toward finance can spread its benefits widely, while suspicion toward financial services can restrict credit, risk-sharing, and competition (Zingales, 2012, 2015). Survey evidence reveals that trust in bankers fell sharply following the 2007–2008 financial crisis (Sapienza and Zingales, 2012), that such public perceptions often diverge from those of economists (Sapienza and Zingales, 2013), and that low trust can hinder insurance market efficiency (Gennaioli, Porta, Lopez-de-Silanes, and Shleifer, 2020). The relatively short time series of survey data restricts our understanding of how finance sentiment changes over time and differs across countries, how it responds to rare disasters like the currently spreading pandemic, and how such changes relate to economic and financial outcomes. While we cannot survey those who lived through the devastating wars and natural disasters of the 20th century, books allow us to travel through time and across borders, and to study public perceptions about the benefits of finance to society.

We measure popular sentiment toward finance in an annual panel covering eight large economies from 1870 to 2009 using a computational linguistics approach applied to the text of millions of books. Our finance sentiment index relies on a recently developed language model (BERT, Devlin, Chang, Lee, and Toutanova, 2018) to measure whether references to finance are, on average, semantically closer to positive versus negative words. BERT and its offsprings have shattered records on multiple natural language processing

tasks, surpassing human ability on many. We use BERT to embed sentences into a relatively low dimensional numerical vector. Following Kozlowski, Taddy, and Evans (2019), we measure the angle between the embedding of sentences mentioning ``finance'' and the ``positive'' minus ``negative'' dimension. This approach goes beyond the dictionary or bag-of-words approach to sentiment analysis (Zhou, 2018) by capturing not only whether a book is positive or negative, but also the degree to which the *context* of the word ``finance'' is positive. By aggregating this positivity angle for all finance-mentioning sentences in each language and in each year, we construct a novel finance sentiment panel.

We find highly persistent differences in finance sentiment across languages. Generally, books written in languages of more capitalist countries tend to discuss finance in a more positive context. Russian finance sentiment is lowest by far throughout our long sample, followed by German, Italian, Chinese, French, and Spanish, with British and American English at the top. Despite considerable within-country variation, this ordering persists throughout our long sample, with the exception of British English finance sentiment, which is slightly higher than American English sentiment until 1912, and slightly lower thereafter. Interestingly, Chinese finance sentiment is about as positive as the French one, though more volatile, temporarily plummeting in 1971 when the People's Republic of China (PRC) is admitted into the United Nations (UN) then rising by a similar amount the following year when US President Nixon visits the PRC, and the Shanghai Communiqué (1972) is issued. Other significant changes in sentiment coincide with major historical events, like wars and revolutions. These major events could affect finance sentiment, but they could also be affected by it or jointly determined by other socio-economic changes. To better understand how the finance sentiment evolves, we analyze how it responds to plausibly exogenous natural disasters (Baker, Bloom, and Terry, 2020).

We document that finance sentiment declines by about 1 percent, one year after a country suffers a severe natural disaster. This average treatment effect, however, hides ample

heterogeneity across disaster types. In particular, epidemics and earthquakes reduce finance sentiment by about 4 percent, while droughts, floods, and landslides increase it by 3, 2 and 5 percent, respectively. The effects of extreme temperature, storms, and smog are statistically no different from zero. These results hold controlling for wars, fatalities, and for country and year fixed effects. Thus, our panel allows us to overcome a common concern about cross-country comparisons that other sources of heterogeneity may be omitted (Guiso, Sapienza, and Zingales, 2004). The inclusion of year fixed effects also means these estimates are not driven by a single common shock such as the 1918 flu pandemic.

What explains these disparate effects? Disaster insurance data, available over the later part of our sample, suggests that epidemic and earthquake risks are largely uninsured by insurance companies, while extreme temperatures, floods, wildfires, and storms are relatively well covered by insurance. Thus, one potential answer, is that finance facilitates risk sharing of some types of risks through insurance, securitization or derivatives, but financial contracts and intermediaries are often designed to prevent ex-post renegotiation (Diamond and Rajan, 2001; Agarwal, Amromin, Ben-David, Chomsisengphet, Piskorski, and Seru, 2017). When insured disasters hit, economic costs are shared broadly across households and generations. But as the COVID-19 pandemic illustrates, when uninsured disasters strike (Walsh, 2020), their damage can be concentrated in parts of the population (Mongey, Pilossoph, and Weinberg, 2020), destroy fragile businesses (Chetty, Friedman, Hendren, and Stepner, 2020), and generate resentment against financial intermediaries (Scism, 2020). Another explanation is that insurance claim disputes affect finance sentiment. Gennaioli, Porta, Lopez-de-Silanes, and Shleifer (2020) show that insurance claims are frequently disputed and result in rejections or lower payments. Sentiment toward insurers may worsen if households learn they are uninsured only after disaster strikes.

The above findings beg the question, does finance sentiment matter? To answer it, we estimate impulse responses to finance sentiment shocks on GDP growth using local

projections (Jordà, 2005). We study GDP growth, an imperfect measure of economic well being, simply because it is available for all countries in our panel. We find that a 1 percent improvement in finance sentiment leads to a gradual and persistent increase in GDP growth of about 20 basis points in each of the ten years following the shock. For a subset of countries that excludes China and Russia, we can include credit growth in the local projections. We find that some but not all of the positive effects of finance sentiment on GDP can be attributed to its positive effect on credit growth.

What do these results imply for the currently spreading COVID-19 pandemic? Assuming its death toll is no greater than the 1918 flu pandemic, our estimates predict a 4 percent contraction in finance sentiment after one year. Such a shock would exacerbate the direct negative effect of the health crisis on economic growth and further reduce cumulative GDP and credit growth by 4 and 5 percentage points, respectively, over the next 5 years.

Our paper relates to recent work on the measurement of public attitude toward the financial sector. Stulz and Williamson (2003) find that a country's language and religion predict its creditor rights. Guiso, Sapienza, and Zingales (2008) find that a general lack of trust reduces stock market participation. Giannetti and Wang (2016) document that after the revelation of corporate fraud in a state, household participation, and trust in the stock market decreases. D'Acunto, Prokopczuk, and Weber (2019) find that present-day demand for finance is lower in German counties where historical antisemitism (and therefore distrust in finance) was higher. Gurun, Stoffman, and Yonker (2018) find that communities indirectly exposed to a Ponzi scheme withdraw assets from investment advisers. Levine, Lin, and Xie (2019) link the African slave trade to household demand and trust of financial services. We contribute to this work by providing a novel measure of sentiment toward finance that spans over a century and several large economies, and documenting how finance sentiment is shaped by natural disasters.

Natural disasters and their effects on the economy are of great interest since the onset of the COVID-19 pandemic. Eisensee and Stromberg (2007) study disaster relief and news coverage. Baker, Bloom, and Terry (2020) use natural disasters as instruments for stock market uncertainty. Jordà, Singh, and Taylor (2020) document persistent declines in real rates of return and increases in wages after pandemics. Closely related is Aksoy, Eichengreen, and Saka (2020), who find that epidemic exposure in an individual's impressionable years negatively affects their confidence in political institutions and leaders.

A broader related literature considers the measurement of culture and its effects on economic outcomes (Guiso, Sapienza, and Zingales, 2006). Cultural differences can persist for generations (Spolaore and Wacziarg, 2013). Changes in culture, ideas, and in particular language, have been tied to the dramatic enrichment the world experienced starting in the 19th century (Mokyr, 2016; McCloskey, 2016). It remains unclear, however, exactly why cultural changes occur (Guiso, Sapienza, and Zingales, 2015). Our results about natural disasters provide one plausibly exogenous cause for such cultural changes.

A recent increase in the availability of textual data has prompted great interest in its use for analysis of culture in particular (Michel, Shen, Aiden, Veres, Gray, Pickett, Hoiberg, Clancy, Norvig, Orwant, et al., 2011), and in economics and finance more broadly (Gentzkow, Kelly, and Taddy, 2019; Loughran and McDonald, 2020). While this literature has yet to study sentiment toward finance or any particular sector, textual analysis has been used to analyze partisanship (Gentzkow and Shapiro, 2010; Luo, Manconi, and Massa, 2020; Goldman, Gupta, and Israelsen, 2020; Engelberg, Henriksson, Manela, and Williams, 2019), product markets (Hoberg and Phillips, 2016), central bank communication (Hansen, McMahon, and Prat, 2018; Cieslak and Vissing-Jorgensen, forthcoming), corporate culture (Grennan, 2019), asset market sentiment (Antweiler and Frank, 2004; Tetlock, 2007; García, 2013; Soo, 2018; Ke, Kelly, and Xiu, 2019), employee expectations (Sheng, 2019), financial constraints (Bodnaruk, Loughran, and McDonald, 2015), subjec-

tive wellbeing (Hills, Proto, Sgroi, and Seresinhe, 2019), uncertainty (Baker, Bloom, and Davis, 2016; Manela and Moreira, 2017; Goetzman, Kim, and Shiller, 2017; Hassan, Hollander, van Lent, and Tahoun, 2017; Boudoukh, Feldman, Kogan, and Richardson, 2018), and emerging risks (Hanley and Hoberg, 2019; Bybee, Kelly, Manela, and Xiu, 2019).

While early work relied on simple word counts (the bag-of-words approach), recent work starting with Mikolov, Chen, Corrado, and Dean (2013) shows that using neural networks to embed words in vector spaces improves learning algorithms' performance in natural language processing tasks. Kozlowski, Taddy, and Evans (2019) demonstrate that such word embeddings produce richer insights into cultural associations and categories than prior methods. Our work builds and improves on their methodology by using a pre-trained language model designed to capture context (BERT), both to embed sentences mentioning our object of interest (finance) and to define the dimension on which we project these embeddings (positive − negative). This ``transfer learning'' approach lowers both estimation error and computation costs.

We proceed as follows. Section 2 describes our text-based finance sentiment measure and how it evolves over time and across countries. Section 3 studies how natural disasters affect finance sentiment. Section 4 analyzes how finance sentiment relates to economic growth. Section 5 concludes. Additional results are provided in an online appendix.

2 Text-based sentiment toward finance

In this section, we describe our text data and how we measure a language's sentiment toward finance across time. For each language and year, we start with a sample of finance-mentioning sentences published in the language and year. Next, we measure the degree to which each sentence places finance in a positive context. We then aggregate these scores to an average finance sentiment that reflects the mean sentiment toward finance of books written in the language in that year.

We assume throughout that the choice of words used by book authors, magazine publishers, and journalists whose work is archived in libraries, reflects the sentiment of the average denizen of that language during the time, or at least that of an influential literary elite. For example, in our dataset, the sentence "correcting corruption or financial malpractice" appears first in 1951 and then appears every year after 1959. The sentence was part of the 1959 US labor management reform legislation hearings, when correcting corruption or financial malpractice became an allowable purpose for establishing a trusteeship by labor unions. Hence, ``financial malpractice'' is more frequently used in subsequent legal documents and books. The context for the word ``financial'' here is clearly negative. In this particular case, we assume that the labor unions in particular, and the US English-speaking public in general, are more likely to associate finance with malpractice around that time.

2.1 Data

Our text data includes five-word sentences (5-grams) containing the word "finance" across eight languages, during 1870–2009, extracted from the 2012 edition of the Google Books Ngram Corpus (Michel, Shen, Aiden, Veres, Gray, Pickett, Hoiberg, Clancy, Norvig, Orwant, et al., 2011; Lin, Michel, Aiden Lieberman, Orwant, Brockman, and Petrov, 2012). The corpus consists of words and phrases and their annual usage frequency from 1500 to 2009. The data originates from Google scanning over 8 million books or 6% of all books ever published in American English, British English, Simplified Chinese, French, German, Italian, Russian, and Spanish.[1]

Although the original data provides lower complexity n-grams counts as well, we focus on 5-grams because for sentiment analysis, especially with BERT, a word's context is essential. We start our study in 1870 (Google corpora is available from 1500) because

[1]We reluctantly omit Hebrew because its word for finance (Mimun) without niqqud is also the name of Maimonides---a famous Jewish philosopher (Rabbi Moshe ben Maimon).

from that year, we have more confidence in the accuracy of our macro data. Moreover, the number of sentences becomes sparser as we go back in time, and there are fewer mentions of finance before 1870, which increases the measurement error of our sentiment index.

Table 1: Finance mentions across languages

Language	Finance word stem	Unique sentences	Total sentences
American English	financ	220k	79m
British English	financ	48k	15m
Simplified Chinese	金融, 金 融, 金_融	196k	305m
French	financ	100k	43m
German	finanz	28k	7m
Italian	finanz	23k	9m
Russian	финан	187k	250m
Spanish	finan	89k	33m

Note: We report the number of mentions of the word finance, translated and stemmed, in a five-word sequence (5-gram) for each language in the Google Book Ngram Corpus. Our dataset covers the period 1870–2009.

We preprocess the Book Corpus by stripping case, symbols, double spaces, part of speech tags, and positional tags. Next, we extract all sentences mentioning the stem of the word for finance. The finance stem word is different across languages, as listed in Table 1. We use the word stem "financ" for English to include sentences that contain either "finance" or "financial." Similarly, for other languages, we use a word stem common to the different verb and noun forms of "finance." For example, for Simplified Chinese we use "金融" (financial) but also include base words where there is space and underscore between 金 (gold) and 融 (melt). The filtering yields a set of unique sentences mentioning finance for each language. In our data set, American English has the highest number of unique sentences that mention "financ", followed by Simplified Chinese and Russian. Although the simplified Chinese is issued starting from the 1950s, the Google ngram data for the Chinese has been translated to the simplified version throughout the whole dataset.

2.2 Methods

We measure finance sentiment across languages at an annual frequency. We employ a three-step process to measure the finance sentiment (i) embed each sentence in our corpus into a 768-dimensional vector space (ii) measure the cosine similarity of this sentence embedding with respect to positive minus negative embedding (iii) average the cosine similarity of all finance mentioning sentences in each year, weighted by their frequency. We next describe how we calculate the sentence embedding, the positive minus negative embedding, and their cosine similarity.

2.2.1 BERT

Recent work in natural language processing (NLP) has been increasingly successful in capturing the complexity of language by considering words in sequence rather than in isolation. One of the ways this is accomplished is by representing words as embeddings. Word embeddings are high-dimensional vector-space models of text in which each unique word in the corpus is represented as a vector in a shared vector space (Mikolov, Chen, Corrado, and Dean, 2013). The vector for each word is based on the context the word shares with other words in the sentence. The classic flavors of word embeddings, such as Word2Vec (Mikolov, Chen, Corrado, and Dean, 2015), GloVe (Pennington, Socher, and Manning, 2014), and FastText (Bojanowski, Grave, Joulin, and Mikolov, 2016; Joulin, Grave, Bojanowski, and Mikolov, 2016) rely on the Distributional Hypothesis (Harris, 1954) to capture relationships in the embedding space. The hypothesis states that words that occur in the same contexts tend to have similar meanings, with the underlying idea that "a word is characterized by the company it keeps" popularized by Firth (1957). However, there are certain downsides with these flavors. First, the traditional methods assign embeddings from the ground up; this is an issue for our data set in earlier years, where the number of words in corpus is less than a million (Altszyler, Sigman, Ribeiro, and Slezak, 2017).

Second, while these embedding methods work well for word-level embeddings, they are poor sentence encoders, which extend the word embedding approach to sentences. (Cer, Yang, Kong, Hua, Limtiaco, John, Constant, Guajardo-Cespedes, Yuan, Tar, Sung, Strope, and Kurzweil, 2018; Perone, Silveira, and Paula, 2018). Thus, we move away from the traditional shallow neural network methods.

We employ a deep neural network-based natural language processing method, Bidirectional Encoder Representations from Transformers (BERT) developed by Devlin, Chang, Lee, and Toutanova (2018). BERT produces meaningful results even with smaller training data and can provide context for words in sentences. The key advantage of this method over classic word vector models is transfer learning – where a model developed for a task is reused as the starting point for a model on a second task. BERT's neural network is pre-trained on 800 million BooksCorpus and 2,500 million Wikipedia words. Thus, the model knows which words have a similar meaning, based on pre-training. BERT is a state of the art NLP model, and Google applies it to both rankings and featured snippets in search. BERT is expected to improve around 10% US English search queries currently, and Google is bringing it to other languages soon.[2] BERT surpassed human performance on the reading comprehension questions provided by the Stanford Question Answering Dataset (SQuAD).

While a full treatment of BERT is beyond our scope, we wish to provide an intuitive understanding of this method and the structure that it implicitly imposes on the data. BERT uses Transformers (Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, and Polosukhin, 2017), a mechanism that learns contextual relations between words in a text. The model processes each word in relation to all other words in a sentence, rather than one-by-one in order. BERT is also bidirectional, which allows the model to learn the context of a word based on all its surroundings, as opposed to a directional model, which reads the text sequentially. To train the model from unlabeled text from BooksCorpus and Wikipedia

[2] https://www.blog.google/products/search/search-language-understanding-bert/

text, BERT employs two strategies: (i) Masked Language Modeling – where 15% of the input words are masked out and then predicted (ii) Next Sentence Prediction – predict if Sentence B is the actual sentence that proceeds Sentence A. Solving the above two problems using its large corpora, BERT is able to place words in the embedding space. Google shares two versions of the pre-trained model: Base (12-layer, 768-hidden features) and Large (24-layer, 1024-hidden features). Both models are available in a cased and uncased variant. We use the base uncased model for English, and Chinese since the extra efficiency we get from the large and cased model is not significant enough to spend more time and resources on them. For French, German, Italian, Russian, and Spanish, we use the cased multilingual model as recommended by Google Research. Thus, we use BERT Base for American and British English, BERT Base Chinese for Simplified Chinese and BERT Base Multilingual Cased for French, German, Italian, Russian and Spanish.[3]

The following features of BERT make it especially useful for our purposes: First, BERT comes pre-trained, so it works well out of the box. A pre-trained model is important to us, especially in earlier years of our sample, where the Google Books corpus is considerably smaller. Second, it offers contextualized embedding. For example, the word "bank" has a different meaning in the following two sentences "In a crisis, we could bank on financing from the government", and "Government's financing for the bank is in crisis." In the first, it means "to rely upon," while in the second, it refers to a financial institution. The context changes how the author feels about the situation. BERT could distinguish the connotation difference between the two sentences, resulting in different embeddings. By contrast, in classical word vector models, where each word has a unique embedding. Third, to reduce the number of unique words that feature in the model, BERT breaks each word into smaller subwords or tokens. For example, "wonderful" is tokenized to "won #der ##ful," where # denotes subwords. The dimensionality reduction is especially important in the multilingual model. Finally, BERT is designed to encode entire sentences, up to 512 subwords.

[3]The pre-trained models are available at https://github.com/google-research/bert

The tokenization process adds [CLS], which stands for "classification" at the beginning of each sentence. The embedding for [CLS] is used as the embedding for the entire sentence that follows it.

2.2.2 Cosine Similarity

A major advantage of word embeddings is that they allow language features (such as words, sentences, etc.) to be treated like vector spaces with intuitive mathematical properties. A common example from Mikolov, Yih, and Zweig (2013) is king – man + woman ~ queen. That is, subtracting the male gender vector and adding the female gender vector to the king vector corresponds to a vector that is close to the queen vector. Thus the word queen could be seen as starting at the word king and then moving in the feminine gender direction. Similarly, we could think of dictator + positive - negative ~ king; here, positive minus negative represent a displacement in the positive direction. Thus, if we start from the dictator vector and move a step in the positive direction, we get the king vector. Other word pairs also correspond to the positive dimension, such as (benefit − damage), (good − bad), (good − corrupt), and (help − hurt).

To define our positive minus negative dimension, we average the sentence embedding differences across sentences containing ``finance'' or ``financial'' together with the above words, similar to Kozlowski, Taddy, and Evans (2019). The list of sentences for English (both American and British) are shown in Table 2. The corresponding sentence pairs for other languages are included in Appendix A.1. We focus on the broader notion of ``finance'', as opposed to more specific financial activities or players (e.g. ``bank'', ``lender'', etc.), because this sentiment measure speaks directly to our question of interest, attitude toward finance. More specific related words would be close in vector space to ``finance'' because they are frequently mentioned together, so we expect them to generate similar sentiment estimates, but each brings along its own identification issues. For example,

``bank turmoil'' can often refer to the financial institution but also to the contested West Bank territory.

Table 2: Positive − negative defining sentences for English

Positive sentences	Negative sentences
financial services benefit society	financial services damage society
finance is good for society	finance is bad for society
finance professionals are mostly good people	finance professionals are mostly corrupt people
finance positively impacts our world	finance negatively impacts our world
financial system helps the economy	financial system hurts the economy

Note: To define the positive minus negative dimension, we average the embeddings of positive sentences less that of their negative counterparts.

To measure sentiment toward finance, for each finance-mentioning sentence j in language i with embedding s_{ji}, we calculate the orthogonal projection of the sentence vector onto the language-specific positivity embedding p_i using cosine similarity:

$$a_{ji} = \frac{s_{ji} \cdot p_i}{|s_{ji}||p_i|} = \frac{\sum_d s_{jid} p_{id}}{\sqrt{\sum_d s_{jid}^2} \sqrt{\sum_d p_{id}^2}}, \tag{1}$$

where d enumerates the elements of s_{ji} and p_i, both 768-dimension vectors. By construction, the cosine similarity in Equation 1 of two positive vectors is bounded between -1 and +1, with zero indicating a neutral sentence. A more negative cosine similarity indicates that the sentence has a more negative sentiment, while a more positive cosine similarity indicates a more positive sentence.

Figure 1 illustrates this method in a two-dimensional space. The five positive (and negative sentences), from Table 2 bunch together in the embedding space, as similar sentences keep similar companies Firth (1957). We take the vector difference between positive and negative sentence embeddings to define our positivity dimension. Next, we project finance sentences onto the positive minus negative dimension. Sentences tend to be close to the dimension, which is closer to their connotation. For example, a sentence such as "financial sector supports economic development" lies closer to the positive sentences, at

a smaller angle with the positive dimension.

Cosine similarity measures the position between -1 and 1 where the shadow of a given sentence vector falls. If the sentence has a positive connotation, such as the one in our example, we will have a smaller angle between the sentence vector and the positive dimension. A smaller angle is associated with a higher cosine similarity. On the other hand, for a negative sentence such as "financial malpractices stunted our growth" would be closer to the negative dimension, or $\theta_{ij} > 90°$. Thus the cosine similarity for a sentence with negative connotation is negative. Similarly, a neutral sentence such as "finance lessons from the pandemic" would be equidistant from both positive and negative dimensions ($\theta_{ij} \approx 90°$), and thus a cosine similarity close to zero. Table 3 lists the sentences with the most positive and most negative finance sentiment for American English. Appendix A.2 provides similar lists for all languages.

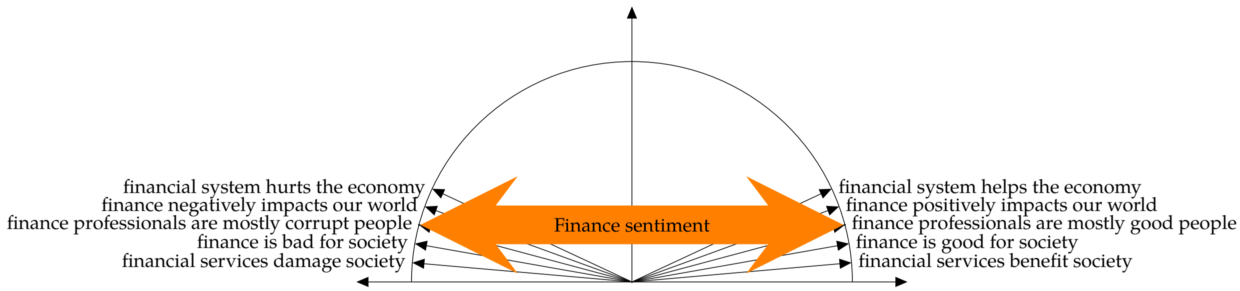
Table 3: Sentences assigned the most positive and negative finance sentiment for English

Positive sentiment sentences	Negative sentiment sentences
financial support of the science	turmoil in the financial markets
financial management of the school	instability in the financial markets
financial support of the research	lack of money to finance
financial management of the business	a financial panic
financial support of this project	the financial panic
financial management initiative	financial panic in the united
financial support of the work	international financial instability
understanding of the financial system	lack of funds to finance
finance for small and medium	my finances falling short
finance graduate school of	the financial deficit

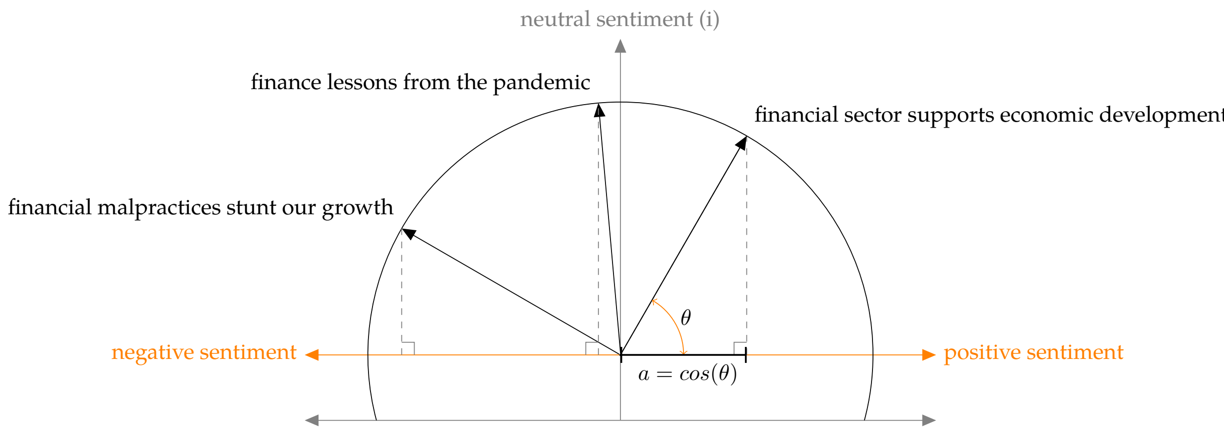
Note: A sentence is assigned positive or negative finance sentiment, based on its projection onto the finance positivity dimension (cosine similarity). Sentences at the top are the most positive or negative in their respective column, and the absolute value of finance sentiment decreases down each list.

Finally, we calculate an annual finance sentiment for each language i based on the cosine similarity of all finance-mentioning sentences that occurs in that language in each

Figure 1: Conceptual diagram of finance sentiment measurement



(a) Defining the positive minus negative finance sentiment dimension



(b) Projection of sentences onto positive minus negative sentiment dimension

Note: Panel (a) shows a conceptual diagram of how similar sentences aggregate to a two-dimensional embedding space. We take the vector difference between positive and negative embeddings to define the finance sentiment dimension. Panel (b) illustrates the classification of three example sentences by projecting them onto this dimension. For one of the sentences, we illustrate cosine similarity, defined as the cosine of the angle between two vectors. Sentences that are close in terms of meaning have a smaller angle between them in this vector space, thus higher cosine similarity. Positive finance sentences have a smaller angle to the positive dimension and a larger positive projection on the finance sentiment dimension.

year t, weighted by the number of times the sentence occurred that year,

$$f_{it} = \sum_j a_{ji} \times \frac{c_{jit}}{\sum_k c_{kit}}. \tag{2}$$

The frequency weighted f_{it} varies over time only because of changes in sentence occurrence c_{jit}, because the sentiment of particular sentences a_{ji} in each language i, does not vary over time. This is an important distinction from the approach of Kozlowski, Taddy,

and Evans (2019), who train a language model for each language in each year, and then measure the orthogonal projections based on these year-specific models. While their approach may be more robust when very large amounts of data are available throughout the sample, our approach is more efficient and avoids issues with measurement error in small samples, which are particularly acute early in the Google Books corpus. Computationally, our approach is considerably cheaper, because training neural networks like the one behind BERT is still fairly expensive.

The cost of this reduction in measurement error and computation cost is that we implicitly assume that the language model is constant over time and that only the frequency of language use varies over time. This is obviously not exactly right. Languages evolve. We lack the data to measure the extent to which such changes in language matter for our conclusions. Encouragingly, evidence from US newspapers over a similar period suggests that changes to the English language do not affect much the ability to predict with text (Manela and Moreira, 2017).

We calculate finance sentiment for every year from 1870 to 2009 for American English, British English, French, German, Italian, Russian, and Spanish. We have a shorter 95-year sample for Chinese because prior to 1922, its corpus is highly sparse and most years feature no mentions of finance. These languages can all be traced to a major geographical area, centered around a distinct country, throughout most of our sample. For example, the concentration of Russian speakers is highest in Russia. Therefore, in what follows, we refer to the finance sentiment of these languages and countries interchangeably, but note that this requires a modest leap of faith. We expect it to introduce more error into our measurement toward the end of our sample, when Spanish books, for example, may be published in Latin American countries whose economic condition is no longer highly correlated with that of Spain.

Another caveat to our finance sentiment index is that it is based on published books,

which may not represent the average citizen, especially early in our sample, when large parts of the world were illiterate. As a result, we may miss marginalized and under-represented groups of the population. Nonetheless, this ``literary elite'' has historically commanded a disproportionately large share of wealth, power, and exerted considerable influence on the opinions of the rest of society.

2.3 Finance sentiment over time and across countries

Table 4 describes finance sentiment over our sample. We see that sentiment toward finance in languages spoken in more capitalist countries tends to be above that of communist countries. In our sample, American English, on average, is at the very top followed closely by British English. The next set of languages that follow are Spanish, French, Chinese, and Italian. German and Russian are the two languages with an overall negative connotation for finance. The standard deviation is highest for Russian, followed by Spanish and French.

Figure 2 plots the finance sentiment and shows some salient features. American English has the most positive sentiment towards finance after 1912; before that, it was slightly below British English. The trend across the language is of an improving finance sentiment across time, with a slight dip at the very end in 2007–2008. A possible reason for that could be the great recession, whose impact could be felt across the globe. We see a 5% drop in US finance sentiment in 1874, a year after the Panic of 1873, which triggered economic depression in Europe and North America. A similar decline in finance sentiment is in 1896, after the country's gold reserves had dwindled and saved by JP Morgan's, and the Rothschild's gold loan. We see an increase in sentiments in 1885 and 1887, after labor union strikes, which eventually led to the eight-hour workday.[4]

Languages do not seem to cross each other, apart from Chinese, which exhibits significant changes in finance sentiment over time. This volatility is in line with historical events.

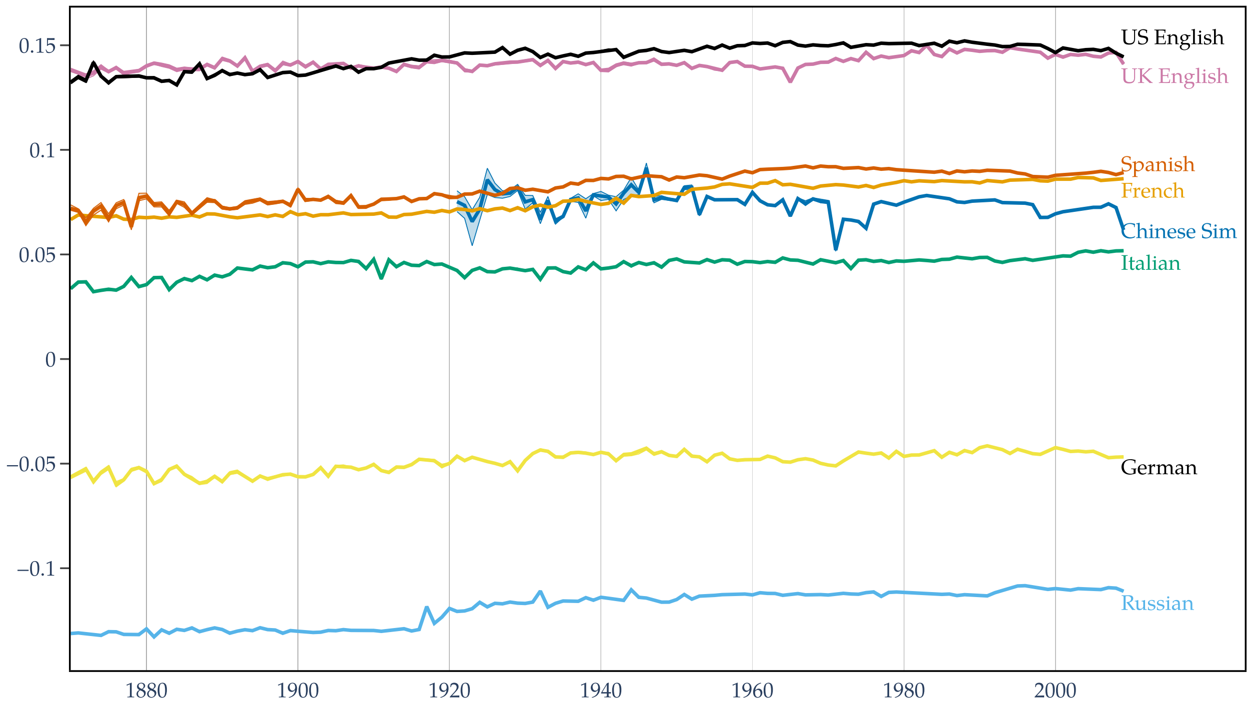
[4]See Wikipedia for historical events mentioned in this section.

Table 4: Finance sentiment and other summary statistics

Country (language)	Variable, %	Mean	Std. Dev.	Obs.
China	Finance sentiment index	7.5	0.5	95
(Chinese)	Finance sentiment growth	0.1	7.9	88
	GDP growth	3.2	7.1	119
France	Finance sentiment index	7.6	0.7	140
(French)	Finance sentiment growth	0.2	1.3	139
	GDP growth	1.9	6.4	139
	Credit growth	4.9	12.9	101
Germany	Finance sentiment index	-4.9	0.5	140
(German)	Finance sentiment growth	0	4.4	139
	GDP growth	2.1	8.1	139
	Credit growth	8.9	17.8	129
Italy	Finance sentiment index	4.4	0.4	140
(Italian)	Finance sentiment growth	0.4	5.2	139
	GDP growth	2	4.7	139
	Credit growth	6.1	13.9	139
Russia	Finance sentiment index	-11.9	0.8	140
(Russian)	Finance sentiment growth	0.1	1.5	139
	GDP growth	2	8.5	139
Spain	Finance sentiment index	8.3	0.7	140
(Spanish)	Finance sentiment growth	0.2	3.4	139
	GDP growth	2.1	5	139
	Credit growth	7.4	11.1	98
UK	Finance sentiment index	14.2	0.3	140
(British English)	Finance sentiment growth	0	1.5	139
	GDP growth	1.5	2.9	139
	Credit growth	4	8.2	129
US	Finance sentiment index	14.5	0.6	140
(American English)	Finance sentiment growth	0.1	1.3	139
	GDP growth	2.1	5	139
	Credit growth	4.5	6.7	129
Total	Finance sentiment index	4.8	8.7	1075
	Finance sentiment growth	0.2	3.7	1061
	GDP growth	2.1	6.2	1092
	Credit growth	5.9	12.5	725

Note: The sample spans from 1870 to 2009 for 8 country-language pairs. The corpus of sentences for each language is gathered from the Google Book Ngram Corpus. The connotation for each finance-mentioning sentence is measured based on its cosine similarity with respect to the positive minus negative vector. Finance sentiment for each year is the weighted average of the cosine similarity, weighted by the frequency of sentences in the language corpus for the year. GDP and credit data are from Jordà, Schularick, and Taylor (2017) and Barro and Ursua (2010) when available.

Figure 2: Sentiment toward finance



Note: Finance sentiment is based on the annual average projection of finance-mentioning sentences' embeddings onto the positive minus negative finance sentiment dimension. Sentences are from the Google Books Ngram corpus and embedded using BERT. Bands represent 95 percent confidence intervals produced by subsampling.

Chinese finance sentiment plunges in 1971 by 31%, just after Mao suggested the end of the Cultural Revolution. However, there is an uptick of 28% in 1972, one year after the United Nations recognized the People's Republic of China as "the only legitimate representative of China", followed by a visit from US President Nixon. We also see an 18% increase in 1976, a year after the constitution of the People's Republic of China was formalized.

The three Romance languages in our sample, French, Italian, and Spanish, have similar attitude towards finance. Spanish has the most favorable view, followed closely by French. We see higher volatility and an uptick in Spanish finance sentiments at the start of the 1874 Bourbon Restoration, which restored the monarchy. French finance sentiment is more volatile during World War II, with the most significant drop of 3% in 1943 when the French surrendered to Germany. The highest surge in French finance sentiment is in 1944, the year Paris was liberated. Sentiment dips and recovers for Italian in 1911–1912 at the start

of the Italo-Turkish war, then rises in 1933 by 14%, when Fascist membership becomes compulsory for University teachers, prompting more favorable and nationalistic literature.

The two languages in which we find a negative finance sentiment are German and Russian. Similar to Italian, we see finance sentiment becoming more positive as the Nazi party gains power in Germany. Finance sentiment increases by 9% in 1930, the year the Nazi party gained its first minister. For Russia, we see a permanent increase in finance sentiment in 1917, coinciding with the Russian revolution. We also see a permanent increase at the beginning of 1990s after the collapse of the USSR, as Russian speaking countries adopt a more capitalist system. The largest drop for Russian finance sentiment is in 1933, a year after the Soviet famine of 1932–1933.

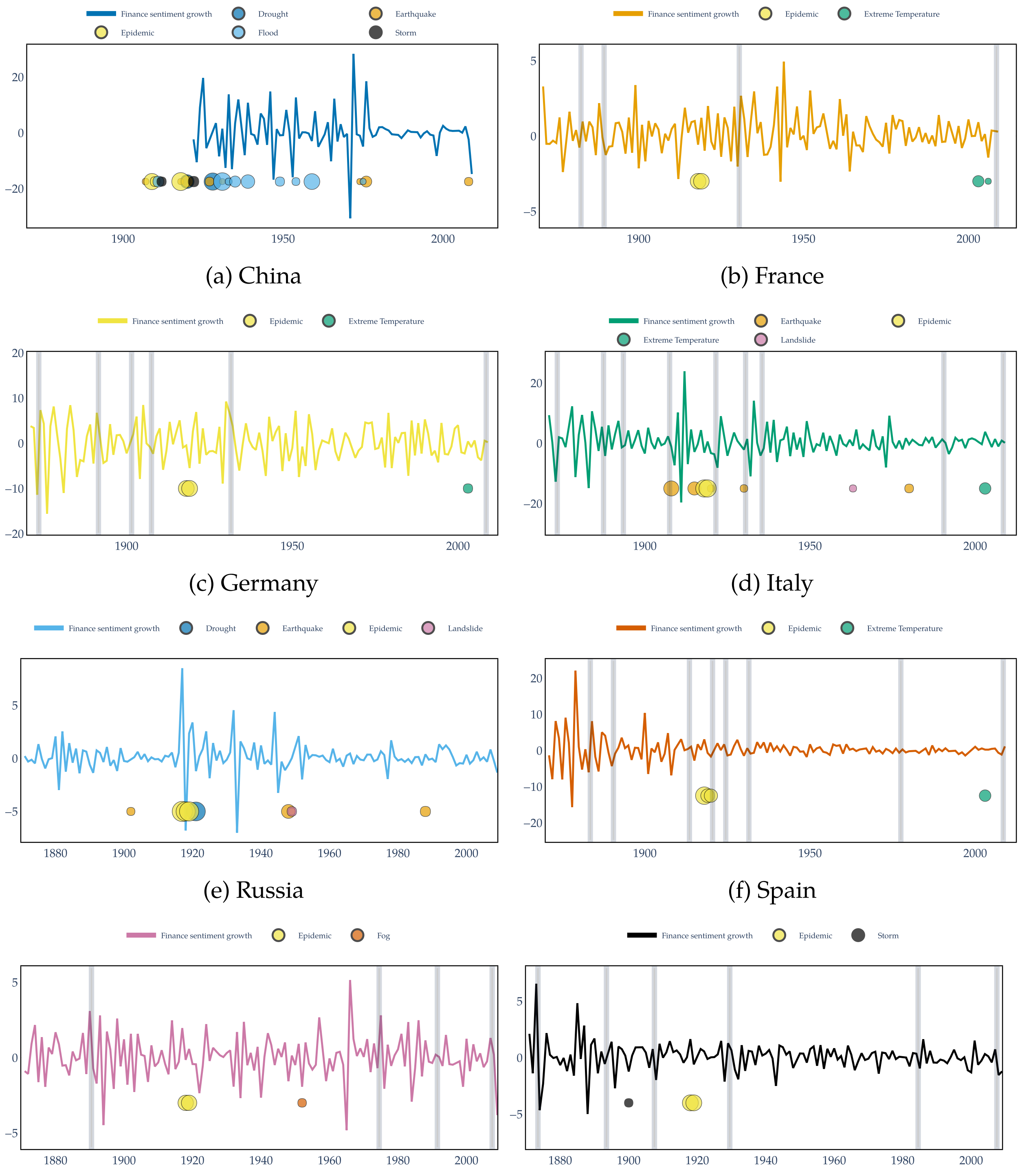
Given the positive trend apparent from Figure 2, our analysis below focuses on finance sentiment growth $\Delta f_{i,t}$, which characterizes the relative change of finance sentiment towards either positive or negative change direction given the absolute value of the previous year's sentiment for country i and year t:

$$\Delta f_{it} = \frac{f_{i,t} - f_{i,t-1}}{|f_{i,t-1}|} \times 100. \tag{3}$$

Figure 3 plots finance sentiment growth for each country in our panel. It shows more clearly that China exhibits the greatest volatility (7.9), followed by Italy (5.2) and Germany (4.4) (see Table 4 for summary statistics). We can also see that sharp changes in finance sentiment growth tend to partially reverse within a year. We formally investigate this pattern in Section 4.

A potential concern may be that it is not just finance-specific sentiment that is changing over time and across countries, but rather sentiment more generally. To measure general sentiment for a language, we use the sentiment associated with the fairly generic word "January" across time, following Gentzkow, Glaeser, and Goldin (2004), who use it to deflate for changes in newspaper reporting volume over time. Appendix Figure 7, shows this

Figure 3: Finance Sentiment Growth



(a) China (b) France (c) Germany (d) Italy (e) Russia (f) Spain (g) UK (h) US

Note: The figure shows finance sentiment growth during the period 1870–2009 for each country. Shades indicate financial crises, as defined by Jordà, Schularick, and Taylor (2017) (not available for China and Russia). Severe natural disasters are indicated by circles whose size is proportional to log deaths.

general sentiment. The general sentiment is quite flat across years, without an upward or a downward trend, for each language. The flat general sentiment suggests that languages have not changed drastically across years. Furthermore, it points to changes in peoples' perception of finance as the underlying cause of the upward trend apparent in Figure 2.

3 Natural disasters affect finance sentiment

We next study how natural disasters and wars affect sentiment toward finance. We first introduce our natural disasters sample and define severe natural disasters. In the next subsection, we present our empirical model consisting of heterogeneity across disaster types and discuss our results.

3.1 Disaster data

Natural disasters data are from the Emergency Events Database (EM-DAT) maintained by the Centre for Research on the Epidemiology of Disasters (CRED). EM-DAT records an event as a disaster if it kills 10 or more people, if it affects 100 or more people, or if there is a formal declaration of a state of emergency or an appeal for international assistance.

To match with our text data, we extract an 8-country subsample of data from EM-DAT dataset, including mainland China, France, Germany, Italy, Russia, Spain, UK and US. Following Eisensee and Stromberg (2007), we focus on natural disasters and omit complex disasters (e.g., famine) and technological disasters (e.g. coal mine collapse), which are likely human-made. We manually add the death tolls caused by the 1918 Flu Pandemic when missing.[5]

[5] The estimated death number from the 1918 Flu Pandemic for each country comes from the academic article and/or the CDC (Center of Disease Control) website or the Spanish flu feature news. Due to the difficulty of accurately determining the real death number caused by the Spanish flu during the period of 1918 to 1919, we distribute the estimated death number equally to each year for those countries where EM-DAT does not specify the death toll.

Table 5: Natural Disasters Summary Statistics

Disaster Group	Type	Obs.	Severe	Mean Killed	Damage, $M	Insured, %	Pub. Lag
Biological	Epidemic	46	19	378133			0.58
Climatological	Drought	20	3	783922	1830		0.00
	Wildfire	53	0	41	504	37.22	
Geophysical	Earthquake	150	18	7534	1744	21.23	0.28
	Volcano	5	0	206	431		
	Mass move.	8	0	79			
Hydrological	Flood	189	9	38949	859	42.97	0.00
	Landslide	66	2	321	224		3.50
Meteorological	Storm	217	3	951	1132	101.20	0.00
	Extreme Temp.	70	5	1068	2233	36.26	0.00
	Fog (Smog)	1	1	4000			0.00
All				35175	1116	83	0.38

Note: Natural disasters by group and type that affect countries in our sample, 1900--2009. For each disaster type, we report the number of disasters, the number of severe disasters, the mean number of people killed, the mean damage (in current USD millions), and the mean percent of damage that is insured. Severe disasters are those that killed at least 20 people per million population. Damage, when available, is the total estimated value of damages and economic losses directly or indirectly related to the disaster in USD millions. Insured losses, when available, are the percent of total damage covered by insurance companies, which sometimes exceed the damage. Publication Lag is the number of years from severe disaster occurrence to its first mention in the corpus.

The EM-DAT dataset classifies disasters by (sub) group and type. Our sample includes 11 distinct natural disaster types, belonging to 5 broader groups. Some countries encounter more than one natural disaster in the same year while other countries experience none. We thus sum the death toll by disaster type within the same year for each country.

Table 5 reports summary statistics for the matched sample of disasters, which includes 825 naturals disasters from 1900 to 2009, 733 of which caused fatalities. While some of these events are clearly salient disasters, some are of a more local nature, and unlikely to change popular sentiment. We therefore, classify a disaster as severe if it kills at least 20per million population, and focus on severe disasters for the most part. As we show below, the exact choice of cutoff is not as important as having a cutoff. The cutoff filters out disasters that affected many people but killed few.

Our analysis thus focuses on 60 severe disasters, of which 32% are epidemics, 30% are earthquakes, and 15% are floods. The table shows that droughts and epidemics were

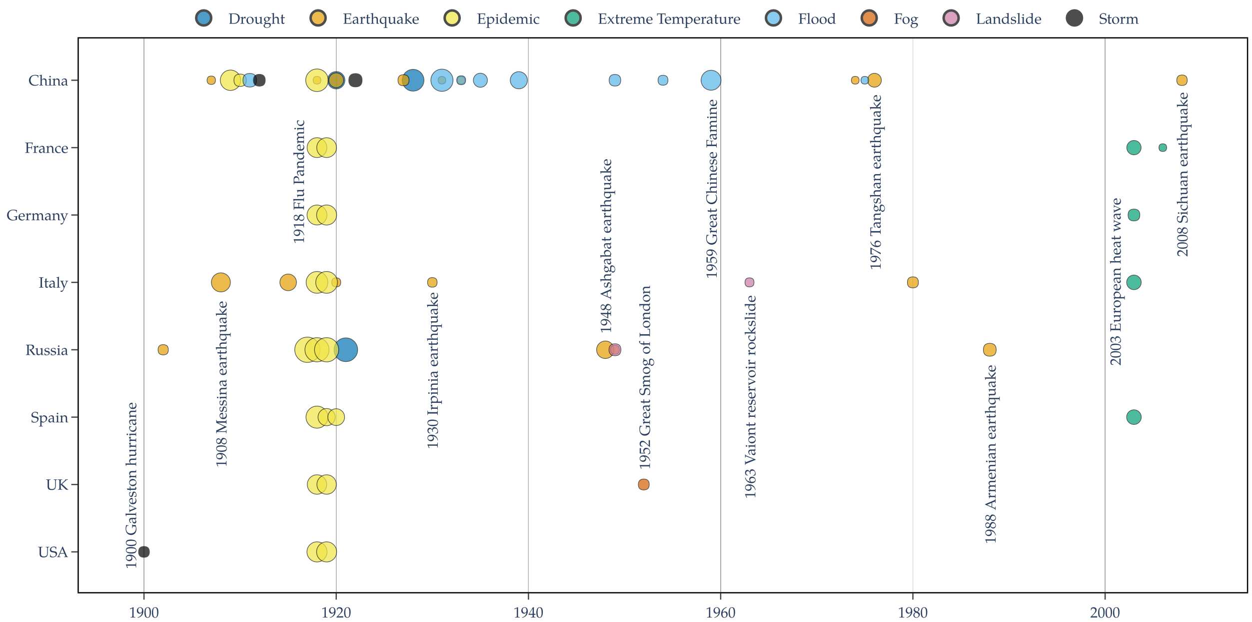
most lethal, killing on average 783 and 378 thousand people. EM-DAT includes an estimate of total damage in current US dollars for some disasters. For a subset of disasters with damage information, EM-DAT reports how much of that damage is covered by insurance companies. Insurance is available for wildfires, earthquakes, floods, storms, and extreme temperature disasters, with sample coverage increasing over time. Floods and storms are well covered by insurance, with storms having higher than 100% coverage. However, earthquakes and especially epidemics are rarely covered by insurance. Since 2017, Munich Re, a large reinsurance company, tried to start underwriting business insurance for epidemics. This effort was mostly unsuccessful until the COVID-19 pandemic hit in late 2019 (Walsh, 2020).

The last column of Table 5 shows that the average severe disaster first appears in the Google Books corpus in the same year that it occurs. This short publication lag is partly the result of books' prominence over most of our sample period as a source of timely information. For example, after a severe hurricane hit the city of Galveston, Texas on September 8, 1900, a book describing the disaster was published in the same year as a fundraising device for the area's devastated public schools (Ousley, 1900). Another reason for this modest average lag is that the Google Books corpus includes many library-stored serial publications. For example, the 1930 Irpinia Earthquake is first mentioned in an information bulletin of the Italian National Research Council (Consiglio nazionale delle ricerche, 1930). Based on this average lag, below we regress finance sentiment growth on one-year lagged disaster indicators.

To investigate the effects of war, we rely on war death tolls from http://necrometrics.com, and concentrate on severe ones as well, which killed a similar fraction of a country's population.

Figure 4 visualizes the distribution of disasters across time and countries. It provides a simple explanation for the lack of insurance coverage for epidemics --- when the 1918 Flu

Figure 4: Severe natural disasters and mortality rates



Note: Circles indicate severe natural disasters, with size proportional to the logarithm of the number of death. We define severe disasters, as disasters with death above 20 per million population.

hit, it spread across the globe within a year and affected all major countries in our sample. Such systematic sources of risk may not provide enough opportunities for risk sharing, and therefore feature high insurance premia and low take up. The figure also shows that severe disasters occur more frequently in developing countries such as China and Russia. UK and US, on the other hand, experienced the 1918 flu and one additional disaster. The dearth of disasters for the more developed economies means that estimates of the effects of natural disasters largely originate in developing counties. Because of the clustering of disasters in time and across countries shown by the figure, we include country and year fixed effects in our analysis below.

3.2 Results

Table 6, Column (1) shows that the mean severe natural disaster hitting a country i in year t decreases next year finance sentiment growth $\Delta f_{i,t+1}$ by about one percentage point. Column (2) considers war as another source of variation in finance sentiment, but shows

that wars do not have a material effect on finance sentiment. Unlike natural disasters, war and its timing, in particular, is endogenous, as it is under the control of the aggressor and partially under the control of the retaliating country, and therefore likely shaped by other economic and political considerations. It is also possible that our war severity indicator is based on realized casualties, but a country's citizens respond more to war news and to expectations of damage (Verdickt, forthcoming). One may expect the number of fatalities caused by a disaster to be more important than the mere occurrence of a natural disaster. However, Column (3) shows that controlling for the number of people killed hardly changes the natural disaster indicator's coefficient or the R-squared. These panel regressions control for country and year fixed effects. Thus unobserved sources of heterogeneity across countries or time, do not confound this result.

The average treatment effect of a severe natural disaster, however, masks considerable heterogeneity. In Column (4), we replace the single disaster indicator, with type-specific disaster indicators that turn on if a disaster of the listed type hits a country in year t. We find that droughts, floods and landslides tend to increase future finance sentiment, while epidemics and earthquakes decrease it significantly. Storms and fog (smog) disasters have large economic effects as well, but cannot be statistically distinguished from zero. Column (5) shows that this result is robust to controlling for disaster fatalities.

The differential effect of a low insurance disaster is considerably negative. As Column (6) shows, the considerable heterogeneity in effects across disaster types can be explained by the variation in insurance coverage mentioned above. To investigate this hypothesis a bit more formally, we define an indicator for low insurance taking the value of 1, if insurance companies covers less than a third of the damage caused by the disaster. Because data on insurance is sparse and concentrated in the latter part of our sample, we impute missing insured percentages for each disaster type, assuming no coverage for missing droughts, epidemics, landslides, volcano eruptions and fogs.[6]

[6]Our conclusions are robust to allowing the cutoff to increase or decrease by 20 percent.

Table 6: Natural disasters affect financial sentiment

	Finance sentiment growth$_{t+1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Natural Disaster$_t$	-0.88**	-0.88**	-0.89**			2.01**
	(0.32)	(0.33)	(0.33)			(0.70)
War$_t$		0.10	0.08			
		(0.40)	(0.42)			
Natural Disaster$_t$ × Low Insured$_t$						-4.44**
						(1.70)
logKilled$_t$			0.10		0.12	
			(0.09)		(0.09)	
Drought$_t$				3.27*	3.60*	
				(1.39)	(1.55)	
Earthquake$_t$				-4.57**	-4.64**	
				(1.88)	(1.92)	
Epidemic$_t$				-4.13**	-4.16**	
				(1.64)	(1.69)	
Extremetemp$_t$				-0.07	-0.05	
				(0.35)	(0.37)	
Flood$_t$				2.39**	2.42***	
				(0.68)	(0.68)	
Landslide$_t$				5.20***	5.41***	
				(1.08)	(1.26)	
Storm$_t$				-5.87	-5.93	
				(4.90)	(5.19)	
Fog$_t$				3.31	3.37	
				(2.57)	(2.50)	
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.13	0.13	0.13	0.16	0.17	0.14
Obs	851	851	851	851	851	851

Note: The dependent variable is finance sentiment percent growth from year t to $t+1$. Natural Disaster and War indicate that a country suffers a severe natural disaster or war in year t, killing at least 20 people per million population. Type-specific indicators for severe disasters are similarly defined. Low insured indicates that the no more than a third of the damage caused by the disaster is covered by insurance. $logKilled$ is the logarithm of the number of deaths plus 1. Standard errors clustered by country are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A potential explanation is the dual roles of finance. Finance facilitates risk sharing through insurance, securitization or derivatives, so that when insured disasters hit, their economic costs are shared broadly, across households and generations. But financial contracts and intermediaries are often designed to prevent ex-post renegotiation (Diamond and Rajan, 2001; Agarwal, Amromin, Ben-David, Chomsisengphet, Piskorski, and Seru, 2017). As a result, damage caused by uninsured disasters can be concentrated in parts of

the population and generate resentment against financial intermediaries.

A related explanation is that sentiment toward financial intermediaries, and insurers in particular, may worsen if households and businesses learn they are uninsured only after the fact. Consistent with this channel, Gennaioli, Porta, Lopez-de-Silanes, and Shleifer (2020) document that insurance claims are frequently disputed, and in countries where this is the norm, insurance policies are more expensive and purchased less.

While the effect heterogeneity between insured and uninsured disasters is intriguing, we note that unlike the disasters themselves, which are plausibly exogenous, unobservable omitted variables may confound the differential effect of insurance. Identifying the exact mechanism is as usual harder than identifying the reduced form effect without an instrument for insurance coverage. For example, insurance markets may be more sophisticated in developed economies or in recent periods due to other technological changes we do not observe.

4 Sentiment and economic growth

We next analyze how finance sentiment growth affects macroeconomic activity. We describe our macroeconomic data, discuss our model specification, and the empirical results.

4.1 Macroeconomic data

The economic and credit data that we use are from the macrohistory dataset compiled by Jordà, Schularick, and Taylor (2017). The macrohistory dataset covers annual data for 17 advanced countries from 1870 to 2016. To merge consistently with our text-based finance sentiment index, we only utilize 6 of them: France, Germany, Italy, Spain, UK, and US, spanning from 1870 to 2009. Together, these 6 countries make up more than 40% of the world economy throughout our sample period. This dataset lacks the GDP and popula-

tion of China and Russia, which we supplement from the Barro-Ursua Macroeconomic Data (Barro and Ursua, 2010). We incorporate credit growth as one of key control variables in our model as credit plays an important role in the macroeconomy and financial development. Following Schularick and Taylor (2012), we use total loans to non-financial private sector as credit proxy.

Table 4 summarizes these macroeconomic variables. GDP growth is highest for China at 3.2%, and all other economies hover around 2%. Germany and Spain exhibit the highest average credit growth.

4.2 Impulse response estimates by local projection

To analyze whether shocks to finance sentiment growth affect economic and credit growth, we estimate cumulative impulse response functions via local projections Jordà (2005):

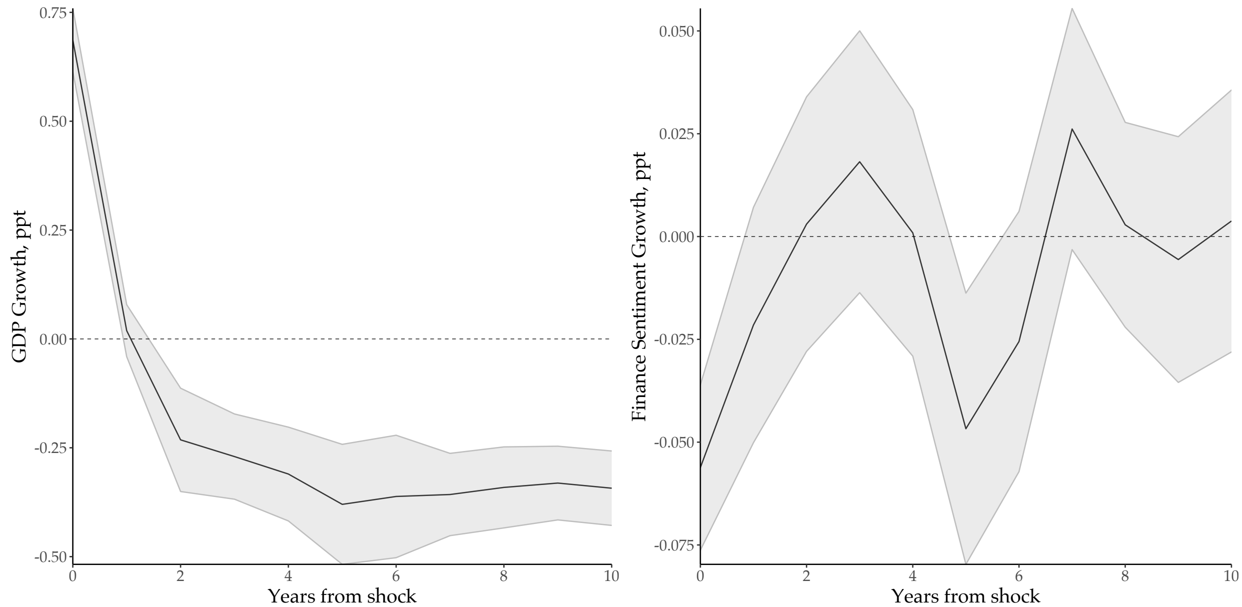
$$\Delta_h y_{i,t+h} = \alpha_i^h + \sum_{k=1}^{3} \beta_k^h \Delta f_{i,t-k} + \sum_{k=1}^{3} \gamma_k^h X_{i,t-k} + \epsilon_{i,t+h}, \quad h = 0, ..., H, \tag{4}$$

where i and t represent the country and year, respectively. α_i^h are country fixed effects, $\Delta_h y_{i,t+h} = y_{i,t+h} - y_{i,t-1}$ indicate the h-year cumulative growth of interest, e.g. GDP growth rate and credit growth, with $h = 0, ..., H$. f_i is finance sentiment index, X_i is vector of control variables, and $\epsilon_{i,t+h}$ are disturbance terms.

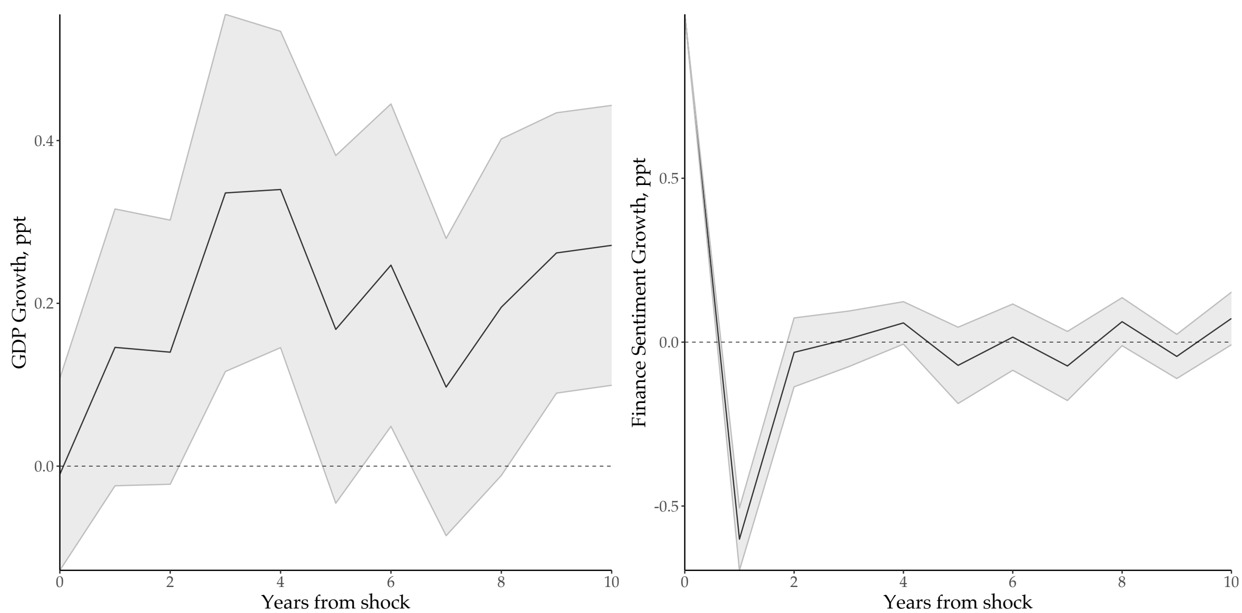
This model estimates the response of $\Delta_h y_{i,t+h}$ from a shock in Δf_i. To capture the direct link of such shocks to economic growth, we control for the first 3 lags of credit growth. Similarly, the first 3 lags of economic growth are control variables when credit growth is our target variable.

The results in Figure 5 include all countries in the sample, and therefore focus on GDP and finance sentiment alone (no credit data is available for China or Russia). The bottom left panel shows that a 1 percentage point increase in finance sentiment growth increases

Figure 5: Impulse response of GDP growth to a finance sentiment growth shock



(a) GDP growth shock



(b) Finance sentiment growth shock

Note: Impulse responses estimated via local projections indicate the change of the cumulative response to a unit shock. Bands are 90% confidence intervals based on Driscoll and Kraay nonparametric robust standard errors.

GDP growth by about 0.3 percentage points 4 years out, though it has no contemporaneous effect. This effect is quite large compared with the mean annual GDP growth of 2.1 percent.

The top right panel shows increases in GDP tend to coincide with declines in finance sentiment growth. While this latter effect is statistically different from zero, its economic magnitude is quite modest.

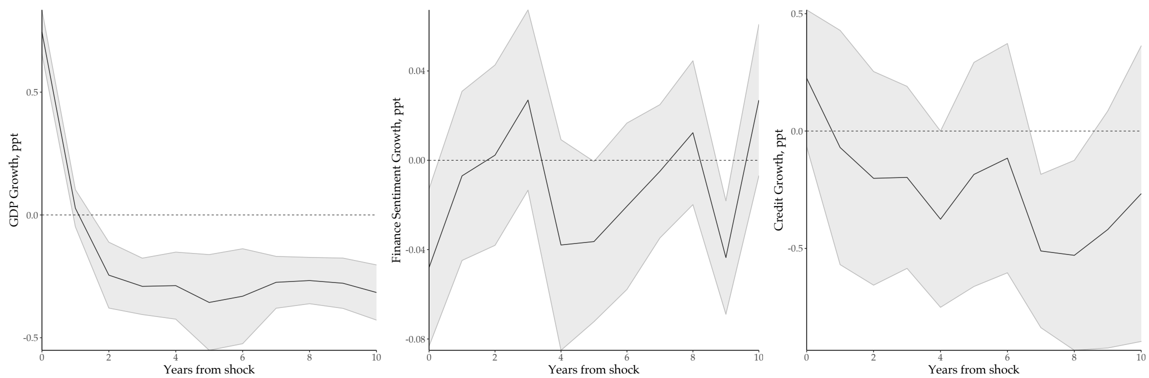
Interestingly, the bottom right panel reveals that finance sentiment growth tends to oscillate after shocks. This is somewhat surprising, as we expected it to gradually mean-revert like GDP growth does on the top right panel. These oscillations could be the result of book writers and publishers attempting to continuously innovate with contrarian books.

It is likely, however, that finance sentiment affects economic growth not directly, but indirectly, by changing the demand for financial services. It may also affect the supply for financial services by changing how the sector is regulated. Both mechanisms should manifest as changes in the quantity of credit. To investigate this channel, we focus next on the subsample of advanced economies for which we have credit data.

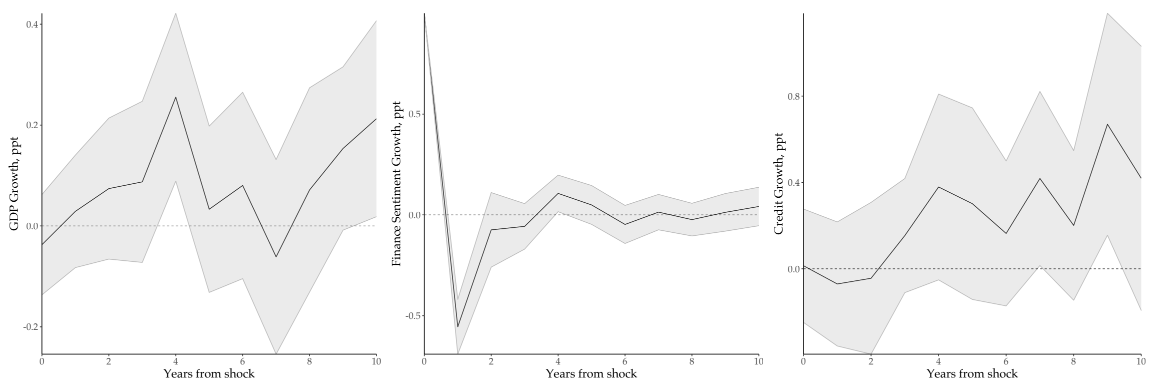
Figure 6 depicts the response of economic growth, finance sentiment growth, and credit growth to shocks by the same three variables. Regardless of the oscillating impact of finance sentiment growth, the cumulative response to the shock in finance sentiment growth is positive for both economic growth and credit growth after a year. A one percentage point increase in finance sentiment growth is associated with a 0.4 percentage point increase in credit growth. The addition of credit growth also reduces somewhat the impulse response of GDP to finance sentiment. It seems, therefore, that some but not all of the effect of finance sentiment on GDP is through credit growth.

What do these estimates imply about the COVID-19 pandemic? From Figure 6, the average effect over the 5 years following the pandemic is a 0.2 percentage point reduction in annual GDP growth, and a 0.25 percentage point reduction in credit growth. Table

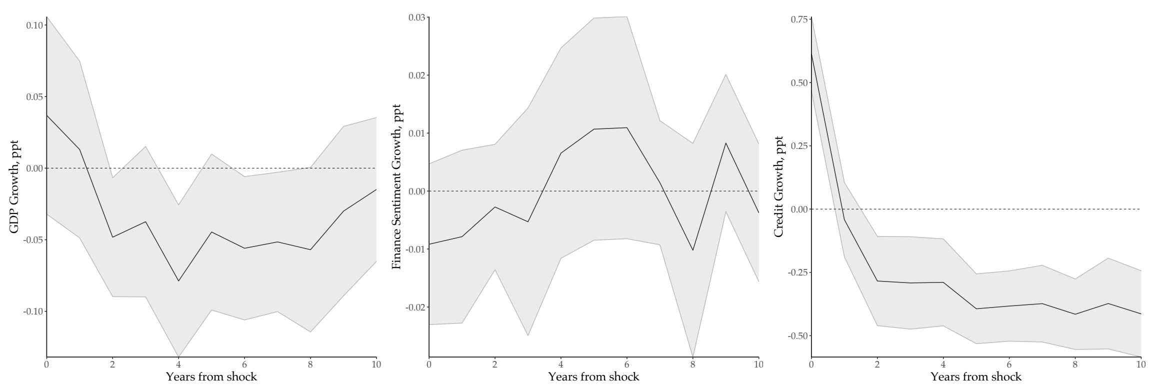
Figure 6: Impulse response of GDP growth and credit growth to finance sentiment growth shocks (without China and Russia)



(a) GDP growth shock



(b) Finance sentiment growth shock



(c) Credit growth shock

Note: Impulse responses estimated via local projections indicate the change of the cumulative response to a unit shock. Bands are 90% confidence intervals based on Driscoll and Kraay nonparametric robust standard errors.

6 shows that the effect of a severe epidemic is a 4 percentage point reduction in finance sentiment growth. Therefore, the cumulative effect of such a decline in finance sentiment on GDP growth over the subsequent 5 years, is about 4 percentage points ($(1.002 \times 4)^5 - 1$). The corresponding effect on credit growth is about 5 percentage points ($(1.0025 \times 4)^5 - 1$).

5 Conclusion and implications for COVID-19

We measure popular sentiment toward finance using a computational linguistics approach applied to millions of books published in eight countries over hundreds of years, and document several new facts.

Finance sentiment differences across countries mostly persist throughout our long sample, with the exception of China, which exhibits great volatility and a level of finance sentiment about as positive as that of Italy and France. Finance sentiment responds negatively to uninsured natural disasters and positively to insured ones. Epidemics and earthquakes, in particular, reduce finance sentiment by about 4 percent within a year. In the VAR sense, shocks to finance sentiment positively affect long term economic and credit growth.

Our estimates imply that beyond its health crisis, the COVID-19 pandemic may reduce GDP growth by 4 percentage points and reduce credit growth by 5 percentage points over the next five years by worsening attitudes toward finance. This back of the envelope calculation assumes, of course, that the COVID-19 pandemic affects finance sentiment like previous severe epidemics. Governments and central bank interventions will hopefully alleviate the pandemic's physical and financial damage and reduce any damage to public sentiment toward finance.

References

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, Tomasz Piskorski, and Amit Seru, 2017, Policy Intervention in Debt Renegotiation: Evidence from the Home Affordable Modification Program, *Journal of Political Economy* 125, 654--712.

Aksoy, Cevat Giray, Barry Eichengreen, and Orkun Saka, 2020, The Political Scar of Epidemics, Working Paper 27401 National Bureau of Economic Research.

Altszyler, Edgar, Mariano Sigman, Sidarta Ribeiro, and Diego Fernández Slezak, 2017, Comparative study of LSA vs Word2vec embeddings in small corpora: A case study in dreams database, *Consciousness and Cognition* 56, 178--187.

Antweiler, Werner, and Murray Z. Frank, 2004, Is all that talk just noise? The information content of Internet stock message boards, *Journal of Finance* 59, 1259--1293.

Baker, Scott, Nicholas Bloom, and Stephen Terry, 2020, Using Disasters to Estimate the Impact of Uncertainty, Discussion Paper w27167 National Bureau of Economic Research Cambridge, MA.

Baker, Scott R, Nicholas Bloom, and Steven J Davis, 2016, Measuring economic policy uncertainty, *Quarterly Journal of Economics* 131, 1593--1636.

Barro, Robert J, and Jose F Ursua, 2010, Barro-ursua macroeconomic data, .

Bodnaruk, Andriy, Tim Loughran, and Bill McDonald, 2015, Using 10-K Text to Gauge Financial Constraints, *Journal of Financial and Quantitative Analysis* 50, 623--646.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov, 2016, Enriching Word Vectors with Subword Information, .

Boudoukh, Jacob, Ronen Feldman, Shimon Kogan, and Matthew Richardson, 2018, Information, Trading, and Volatility: Evidence from Firm-Specific News, *Review of Financial Studies*.

Bybee, Leland, Bryan T. Kelly, Asaf Manela, and Dacheng Xiu, 2019, The Structure of Economic News, *SSRN Electronic Journal*.

Cer, Daniel, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil, 2018, Universal Sentence Encoder, *arXiv:1803.11175 [cs]*.

Chen, Yanqing, and Steven Skiena, 2014, Building Sentiment Lexicons for All Major Languages, in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* pp. 383--389 Baltimore, Maryland. Association for Computational Linguistics.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, and Michael Stepner, 2020, Real-Time Economics: A New Platform to Track the Impacts of COVID-19 on People, Businesses, and Communities Using Private Sector Data, Discussion paper Mimeo.

Cieslak, Anna, and Annette Vissing-Jorgensen, forthcoming, The Economics of the Fed Put, *Review of Financial Studies*.

Consiglio nazionale delle ricerche, 1930, *Bollettino d'informazioni* (Consiglio nazionale delle ricerche).

D'Acunto, Francesco, Marcel Prokopczuk, and Michael Weber, 2019, Historical Antisemitism, Ethnic Specialization, and Financial Development, *Review of Economic Studies* 86, 1170--1206.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, 2018, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *arXiv:1810.04805 [cs]*.

Diamond, Douglas W., and Raghuram G. Rajan, 2001, Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking, *Journal of Political Economy* 109, pp. 287--327.

Eisensee, T., and D. Stromberg, 2007, News Droughts, News Floods, and U. S. Disaster Relief, *Quarterly Journal of Economics* 122, 693--728.

Engelberg, Joseph, Matthew Henriksson, Asaf Manela, and Jared Williams, 2019, The Partisanship of Financial Regulators, Working Paper.

Firth, JR, 1957, Applications of General Linguistics - Firth - 1957 - Transactions of the Philological Society - Wiley Online Library, .

García, Diego, 2013, Sentiment during recessions, *Journal of Finance* 68, 1267--1300.

Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer, 2020, Trust and Insurance Contracts, Discussion Paper w27189 National Bureau of Economic Research Cambridge, MA.

Gentzkow, Matthew, Edward L Glaeser, and Claudia Goldin, 2004, The Rise of the Fourth Estate: How Newspapers Became Informative and Why It Mattered, Working Paper 10791 National Bureau of Economic Research.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, 2019, Text as Data, *Journal of Economic Literature* 57, 535--574.

Gentzkow, Matthew, and Jesse M. Shapiro, 2010, What Drives Media Slant? Evidence From U.S. Daily Newspapers, *Econometrica* 78, 35--71.

Giannetti, Mariassunta, and Tracy Yue Wang, 2016, Corporate Scandals and Household Stock Market Participation, *Journal of Finance* 71, 2591--2636.

Goetzman, W., D. Kim, and R. J. Shiller, 2017, Affect, Media and Earthquakes: Determinants of Crash Beliefs from Investor Surveys, Discussion paper Yale University Working Paper.

Goldman, Eitan, Nandini Gupta, and Ryan D. Israelsen, 2020, Political Polarization in Financial News, SSRN Scholarly Paper ID 3537841 Social Science Research Network Rochester, NY.

Grennan, Jillian, 2019, A Corporate Culture Channel: How Increased Shareholder Governance Reduces Firm Value, Working Paper ID 2345384 Rochester, NY.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2004, Does Local Financial Development Matter?, *Quarterly Journal of Economics* 119, 929--969.

———, 2006, Does Culture Affect Economic Outcomes?, *Journal of Economic Perspectives* 20, 23--48.

———, 2008, Trusting the stock market, *Journal of Finance* 63, 2557--2600.

———, 2015, Corporate Culture, Societal Culture, and Institutions, *American Economic Review* 105, 336--339.

Gurun, Umit G., Noah Stoffman, and Scott E. Yonker, 2018, Trust Busting: The Effect of Fraud on Investor Behavior, *Review of Financial Studies* 31, 1341--1376.

Hanley, Kathleen Weiss, and Gerard Hoberg, 2019, Dynamic Interpretation of Emerging Risks in the Financial Sector, *Review of Financial Studies*.

Hansen, Stephen, Michael McMahon, and Andrea Prat, 2018, Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach, *Quarterly Journal of Economics* 133, 801--870.

Harris, Zellig S., 1954, Distributional Structure, *WORD* 10, 146--162.

Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun, 2017, Firm-Level Political Risk: Measurement and Effects, SSRN Scholarly Paper ID 2838644 Social Science Research Network Rochester, NY.

Hills, Thomas T., Eugenio Proto, Daniel Sgroi, and Chanuki Illushka Seresinhe, 2019, Historical analysis of national subjective wellbeing using millions of digitized books, *Nature Human Behaviour* 3, 1271--1275.

Hoberg, Gerard, and Gordon Phillips, 2016, Text-Based Network Industries and Endogenous Product Differentiation, *Journal of Political Economy* 124, 1423--1465.

Jordà, Òscar, 2005, Estimation and Inference of Impulse Responses by Local Projections, *American Economic Review* 95, 161--182.

———, Moritz Schularick, and Alan M. Taylor, 2017, Macrofinancial History and the New Business Cycle Facts, *NBER Macroeconomics Annual* 31, 213--263.

Jordà, Òscar, Sanjay R Singh, and Alan M Taylor, 2020, Longer-run Economic Consequences of Pandemics, Working Paper 26934 National Bureau of Economic Research.

Joulin, Armand, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov, 2016, Bag of Tricks for Efficient Text Classification, *arXiv e-prints* 1607, arXiv:1607.01759.

Ke, Zheng Tracy, Bryan T. Kelly, and Dacheng Xiu, 2019, Predicting Returns with Text Data, Working Paper.

Kozlowski, Austin C., Matt Taddy, and James A. Evans, 2019, The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings:, *American Sociological Review*.

Levine, Ross, Chen Lin, and Wensi Xie, 2019, The African Slave Trade and Modern Household Finance, SSRN Scholarly Paper ID 3031310 Social Science Research Network Rochester, NY.

Lin, Yuri, Jean-Baptiste Michel, Erez Aiden Lieberman, Jon Orwant, Will Brockman, and Slav Petrov, 2012, Syntactic Annotations for the Google Books NGram Corpus, in *Proceedings of the ACL 2012 System Demonstrations* pp. 169--174 Jeju Island, Korea. Association for Computational Linguistics.

Loughran, T., and B. McDonald, 2011, When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *Journal of Finance* 66, 35--65.

Loughran, Tim, and Bill McDonald, 2020, Textual Analysis in Finance, SSRN Scholarly Paper ID 3470272 Social Science Research Network Rochester, NY.

Luo, Mancy, Alberto Manconi, and Massimo Massa, 2020, Blinded by Perception? The Stock Market's Reaction to Politically Aligned Media, SSRN Scholarly Paper ID 2879939 Social Science Research Network Rochester, NY.

Manela, Asaf, and Alan Moreira, 2017, News implied volatility and disaster concerns, *Journal of Financial Economics* 123, 137--162.

McCloskey, Deirdre Nansen, 2016, *Bourgeois Equality: How Ideas, Not Capital or Institutions, Enriched the World* (University of Chicago Press).

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K Gray, Joseph P Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, et al., 2011, Quantitative analysis of culture using millions of digitized books, *Science* 331, 176--182.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean, 2013, Efficient Estimation of Word Representations in Vector Space, *arXiv:1301.3781 [cs]*.

Mikolov, Tomas, Kai Chen, Gregory S. Corrado, and Jeffrey A. Dean, 2015, Computing numeric representations of words in a high-dimensional space, .

Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig, 2013, Linguistic Regularities in Continuous Space Word Representations, in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* pp. 746--751 Atlanta, Georgia. Association for Computational Linguistics.

Mokyr, Joel, 2016, *A Culture of Growth: The Origins of the Modern Economy* (Princeton University Press).

Mongey, Simon, Laura Pilossoph, and Alex Weinberg, 2020, Which Workers Bear the Burden of Social Distancing Policies?, Working Paper 27085 National Bureau of Economic Research.

Ousley, Clarence, 1900, *Galveston in Nineteen Hundred: The Authorized and Official Record of the Proud City of the Southwest as It Was Before and After the Hurricane of September 8, and a Logical Forecast of Its Future* (Brookhaven Press).

Pennington, Jeffrey, Richard Socher, and Christopher Manning, 2014, GloVe: Global Vectors for Word Representation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* pp. 1532--1543 Doha, Qatar. Association for Computational Linguistics.

Perone, Christian S., Roberto Silveira, and Thomas S. Paula, 2018, Evaluation of sentence embeddings in downstream and linguistic probing tasks, *arXiv:1806.06259 [cs]*.

Sapienza, Paola, and Luigi Zingales, 2012, A Trust Crisis, *International Review of Finance* 12, 123--131.

——— , 2013, Economic Experts versus Average Americans, *American Economic Review* 103, 636--642.

Schularick, Moritz, and Alan M Taylor, 2012, Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008, *American Economic Review* 102, 1029--1061.

Scism, Leslie, 2020, Companies Hit by Covid-19 Want Insurance Payouts. Insurers Say No., *Wall Street Journal*.

Sheng, Jinfei, 2019, Asset Pricing in the Information Age: Employee Expectations and Stock Returns, SSRN Scholarly Paper ID 3321275 Social Science Research Network Rochester, NY.

Soo, Cindy K., 2018, Quantifying Sentiment with News Media across Local Housing Markets, *Review of Financial Studies* 31, 3689--3719.

Spolaore, Enrico, and Romain Wacziarg, 2013, How Deep Are the Roots of Economic Development?, *Journal of Economic Literature* 51, 325--369.

Stulz, René M., and Rohan Williamson, 2003, Culture, openness, and finance, *Journal of Financial Economics* 70, 313--349.

Tetlock, Paul C., 2007, Giving Content to Investor Sentiment: The Role of Media in the Stock Market, *Journal of Finance* 62, 1139--1168.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, 2017, Attention Is All You Need, *arXiv:1706.03762 [cs]*.

Verdickt, Gertjan, forthcoming, The Effect of War Risk on Managerial and Investor Behavior: Evidence from the Brussels Stock Exchange in the Pre-1914 Era, *Journal of Economic History*.

Walsh, Mary Williams, 2020, Coronavirus Will Cost Businesses Billions. Insurance May Not Help., *The New York Times*.

Zhou, Guofu, 2018, Measuring Investor Sentiment, *Annual Review of Financial Economics* 10, 239--259.

Zingales, Luigi, 2012, *A Capitalism for the People: Recapturing the Lost Genius of American Prosperity* (Basic Books).

——— , 2015, Presidential Address: Does Finance Benefit Society?, *Journal of Finance* 70, 1327--1363.

A Online Appendix

A.1 Positive and negative sentences used to define the positivity dimension across languages

Table 7: Positive and negative sentences

Positive sentences	Negative sentences
financial services benefit society	financial services damage society
finance is good for society	finance is bad for society
finance professionals are mostly good people	finance professionals are mostly corrupt people
finance positively impacts our world	finance negatively impacts our world
financial system helps the economy	financial system hurts the economy

(a) English

金融服务有益社会	金融服务损害社会
金融对社会好	金融对社会不好
财务专业人员大多很好	财务专业人员大多邪恶
金融对世界产生积极影响	金融对世界产生消极影响
金融系统帮助经济	金融系统有害金融

(b) Chinese

les services financiers profitent à la société	les services financiers nuisent à la société
la finance est bonne pour la société	la finance est mauvaise pour la société
les professionnels de la finance sont surtout bons	les professionnels de la finance sont surtout mauvais
la finance a un impact positif sur notre monde	la finance a un impact négatif notre monde
le système financier aide l'économie	le système financier nuit à l'économie

(c) French

Table 7: Positive and negative sentences, continued.

Positive sentences	Negative sentences
Finanzdienstleistungen kommen der Gesellschaft zugute	Finanzdienstleistungen schaden der Gesellschaft
Finanzen sind gut für die Gesellschaft	Finanzen sind schlecht für die Gesellschaft
Finanzprofis sind meistens gut	Finanzprofis sind meistens böse
Finanzen wirken sich positiv auf unsere Welt aus	Finanzen wirken sich negativ aus unsere Welt
Finanzsystem hilft der Wirtschaft	Finanzsystem schadet der Wirtschaft

(d) German

i servizi finanziari avvantaggiano la società	i servizi finanziari danneggiano la società
la finanza fa bene alla società	la finanza fa male alla società
i professionisti della finanza sono per lo più buoni	i professionisti della finanza sono principalmente cattivi
la finanza ha un impatto positivo sul nostro mondo	la finanza ha un impatto negativo il nostro mondo
il sistema finanziario aiuta l'economia	il sistema finanziario danneggia l'economia

(e) Italian

общество оказывает финансовую помощ	общество наносит ущерб финансовым услугам
финансы полезны для общества	финансы вредны для общества
профессионалы в области финансов в основном хороши	профессионалы в области финансов в основном злые
финансы положительно влияют на наш мир	финансы негативно влияют наш мир
финансовая система помогает экономике	финансовая система наносит ущерб экономике

(f) Russian

los servicios financieros benefician a la sociedad	los servicios financieros perjudican a la sociedad
los profesionales financieros son en su mayoría buenos	los profesionales financieros son en su mayoría malos
las finanzas impactan positivamente en nuestro mundo	las finanzas impactan negativamente nuestro mundo
el sistema financiero ayuda a la economía	el sistema financiero perjudica a la economía

(g) Spanish

Note: In line with Kozlowski, Taddy, and Evans (2019), we start with five pairs of words for the positive minus negative dimension for English (both American and British). The word pair includes: (positive − negative), (benefit − damage), (good − bad), (good − corrupt), and (help − hurt). We then create positive and negative sentences which discuss finance, using these words. For other languages, we translate these sentences with the help of native speakers.

A.2 Top ten worst to best ngrams sorted by finance sentiment

Table 8: Worst to best sentence sorted by finance sentiment

American English	British English
turmoil in the financial markets	turmoil in the financial markets
finances become disordered the	instability in the financial markets
financial panic swept the country	lack of money to finance
turmoil in financial markets	a financial panic
financial panic swept the nation	the financial panic
instability in the financial markets	financial panic in the united
financial panic in the country	international financial instability
severe financial setbacks	lack of funds to finance
a major financial panic	my finances falling short
world wide financial panic	the financial deficit
⋮	⋮
knowledge of the financial structure	finance graduate school of
financial support of the field	finance for small and medium
financial support of the course	understanding of the financial system
financial support of the science	financial support of the work
financial support of the graduate	financial management initiative
business and financial experience	financial support of this project
financial management of the organization	financial management of the business
financial support of the center	financial support of the research
finance in the graduate school	financial management of the school
the goal of financial management	financial support of the science

(a) English

Table 8: Worst to best sentence sorted by finance sentiment, continued

Chinese	English Translation
严重 扰乱 了 金融 秩序	Seriously disturbed the financial order
扰乱 了 国家 金融 秩序	Disrupt the national financial order
严重 扰乱 了 金融	Seriously disrupting the financial
扰乱 了 正常 的 金融	Disrupt the normal financial
扰乱 了 金融 秩 序	Disrupt the financial order of rank
扰乱 了 金融 秩序	Disrupt the financial order
扰乱 了 金融 市场	Disrupt the financial markets
干扰 了 金融 秩序	Disturb financial order
既 不 利于 金融	Not only is not conducive to financial
扰乱 了 金融 序	Disrupt the financial order
⋮	⋮
经济 发展 提供 金融	Economic development has provided financial
农村 发展 提供 金融	Rural Development provides financial
金融 推动 发展	Promote the development of financial
金融 服务 促进 农村	Promotion of rural financial services
金融 务 促进	Promote financial affairs
服务 促进 金融	Promoting financial services
金融 立足	Financial foothold
服务 农村 金融	Financial services in rural areas
金融 服务 社会	Financial services community
服务 规范 发展 金融	Regulate the development of financial services

(b) Chinese

Table 8: Worst to best sentence sorted by finance sentiment, continued

French	English Translation
ny a pas de finances	ny no finances
ministre des finances rené pleven	Finance Minister Rene pleven
the financial revolution in england	the financial revolution in england
état des finances était déplorable	financial condition was deplorable
bérenger finances et absolutisme	bérenger Finance and absolutism
finances est rejeté	Finance is rejected
mauvais état des finances royales	poor state of the royal finances
état des finances na pas	financial condition didnt
finances étaient en mauvais état	finances were in bad condition
finances na pu être déposé	na been filed Finance
⋮	⋮
encourager et à soutenir financièrement	encourage and support financially
mobiliser les ressources financières et	mobilize financial resources and
la gestion financière en	financial management
assistance financière et technique avec	financial and technical assistance with
à la coopération financière avec	financial cooperation with
réaliser la solidarité financière des	achieve financial solidarity
assurer la gestion financière et	the financial management and
organiser et de financer les	organize and finance
de promouvoir et de financer	promote and finance
à promouvoir et à financer	to promote and finance

(c) French

Table 8: Worst to best sentence sorted by finance sentiment, continued

German	English Translation
christian watrin bochum finanzpolitik	christian watrin Bochum financial policy
finanzmarkt kapitalismus	financial market capitalism
renzsch wolfgang finanzverfassung	renzsch wolfgang financial constitution
imperialismus staatsfinanzen rüstung	imperialismus government finances armor
gemeindefinanzgesetz vom dezember	community financial law from december
neoabsolutismus staatsfinanzen und politik	Absolutism government finances and politics
r a finanztheorie	r a financial theory
r a musgrave finanztheorie	r a musgrave finance theory
schmölders finanzpolitik berlin	Schmölders financial policy berlin
mayer geschichte der finanzwirtschaft	mayer history of finance economy
⋮	⋮
ist zuständig für die finanzielle	is responsible for the financial
finanziell und organisatorisch zu unterstützen	financial and organizational support
hilfe bei der finanzierung der	help with the financing of
finanzen die zur durchführung und	finance the implementation and to
hilfe bei der finanzierung von	help with the financing of
finanzieren mit	fund with
finanzierung erfolgt durch beiträge der	financed through contributions of the
unternehmen damit derartige finanzierungen	company so that such financing
sorgt für die finanzierung	provides for the financing
finanzen und mit zustimmung des	Finance and with the approval of

(d) German

Table 8: Worst to best sentence sorted by finance sentiment, continued

Italian	English Translation
dimissioni del ministro delle finanze	resignation of Finance Minister
il finanziamento è stato concesso	The loan was granted
le finanze sono condannate dai	finances are condemned by
scioglimento del contratto di finanziamento	termination of the loan agreement
ministro delle finanze è autorizzato	Minister of Finance is authorized
grave crisi finanziaria	serious financial crisis
il ministro delle finanze dichiarava	Finance Minister declared
le finanze saranno emanate	finances will be issued
la finanza sabauda allaprirsi	finance Savoy allaprirsi
l esercizio finanziario ha inizio	the financial year
⋮	⋮
disponibilità di risorse finanziarie che	availability of financial resources
di gestire le risorse finanziarie	to manage the financial resources
a soddisfare le esigenze finanziarie	to meet the financial needs
effettuare la gestione finanziaria di	make the financial management of
gestione delle risorse finanziarie e	management of financial resources and
relazioni economiche e finanziarie con	economic and financial relations with
coordinamento della finanza regionale con	coordination of regional finance with
assistere tecnicamente e finanziariamente i	assist technically and financially
gestione delle risorse finanziarie idonee	management of the financial resources
relazioni commerciali e finanziarie con	commercial and financial relations with

(e) Italian

Table 8: Worst to best sentence sorted by finance sentiment, continued

Russian	English Translation
обращение финансы кредит	recourse finance loan
плутократия бароны финансового	plutocracy financial barons
буржуазии финансовый срыв	Financial breakdown of the bourgeoisie
протекционизм господство финансистов	Protectionism domination of financiers
финансов кредита социализме	Finance socialism loan
империализм финансовый капитализм	financial capitalism, imperialism
обращение кредит финансы	recourse loan finance
финансовое банкротство	financial bankruptcy
империализма колониальное финансовое порабощение	colonial imperialism financial enslavement
страшных финансовых грозных	terrible financial formidable
⋮	⋮
оказывает и финансовую поддержку	and providing financial support
оказывает колхозам финансовую помощь	It provides financial assistance to collective farms
финансовая деятельность колхоза осуществляется на	financial activities carried out on a collective farm
обеспечивается финансирование мероприятий	provided funding
оказывает финансовую и помощь	and provides financial assistance
оказывает финансовую и поддержку	It provides financial support and
оказывает финансовую и политическую поддержку	It provides financial and political support
оказывает большую финансовую помощь	providing more financial aid
оказывает значительную финансовую помощь	providing substantial financial assistance
оказывает финансовую и техническую помощь	It provides financial and technical assistance

(f) Russian

Table 8: Worst to best sentence sorted by finance sentiment, continued

Spanish	English Translation
la finanza no era	finance was not
la situación financiera no era	the financial situation was not
el capital financiero se sentirá	financial capital will feel
el capital financiero no es	financial capital is not
el mercado financiero no es	the financial market is not
la especulación financiera domine su	financial speculation dominates its
el sistema financiero se vio	the financial system was
una desgraciada situación financiera pudiese	an unfortunate financial situation could
el déficit se financió	The deficit was financed
su situación financiera no era	its financial situation was not
⋮	⋮
actividades financieras y de servicios	financial activities and services
asesoría técnica y apoyo financiero	technical advice and financial support
financiamiento de las diversas actividades	financing various activities
apoyo técnico y financiero internacional	international technical and financial support
apoyo financiero y asistencia técnica	financial support and technical assistance
apoyo financiero a las actividades	financial support to activities
apoyo financiero para las actividades	Financial support for activities
asistencia técnica y recursos financieros	technical assistance and financial resources
financiamiento de las actividades culturales	financing of cultural activities
asistencia técnica y de financiamiento	technical assistance and financing

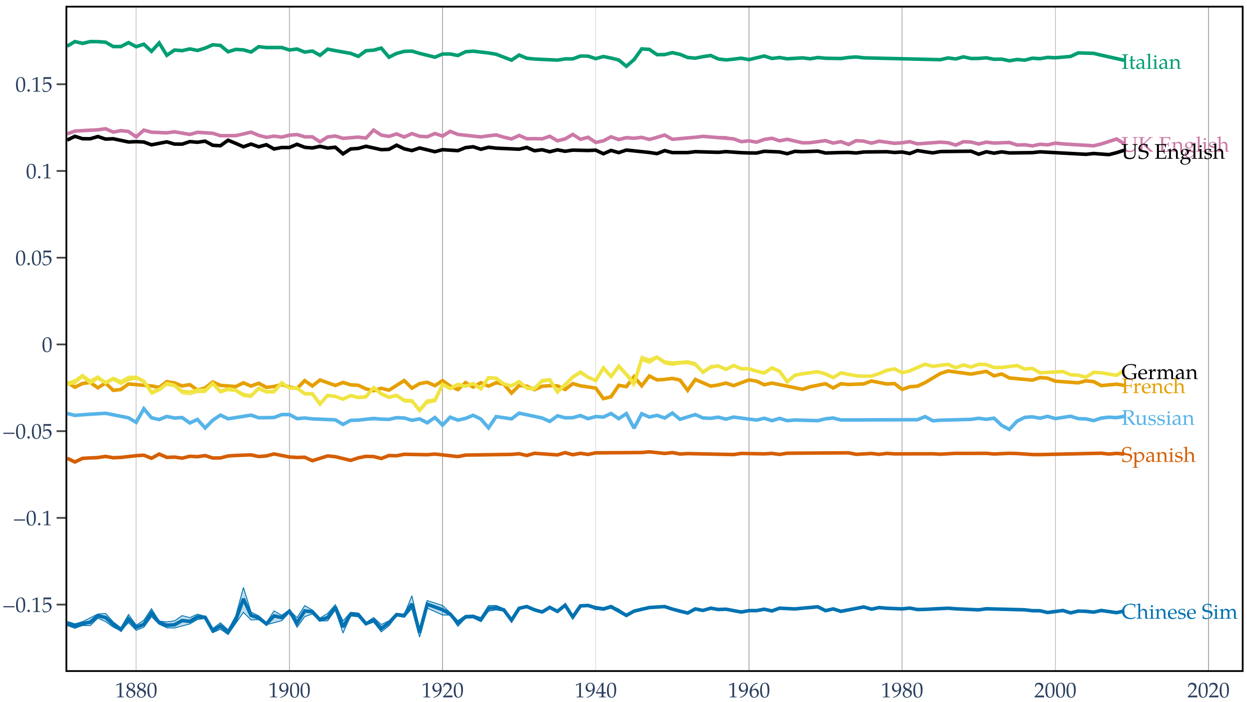
(g) Spanish

Note: The sentences are sorted from worst to best in terms of their cosine similarity with positive minus negative vector. The English translation is provided using Google Translate.

A.3 General sentiment for each language, not just sentences mentioning finance

In Figure 7 we repeat our exercise, but instead of finance-mentioning sentences, we measure general sentiment by focusing on January-mentioning sentences. We find no time trend in general sentiment, and a ranking across languages that is quite different from that of the finance sentiment shown in Figure 2.

Figure 7: General sentiment



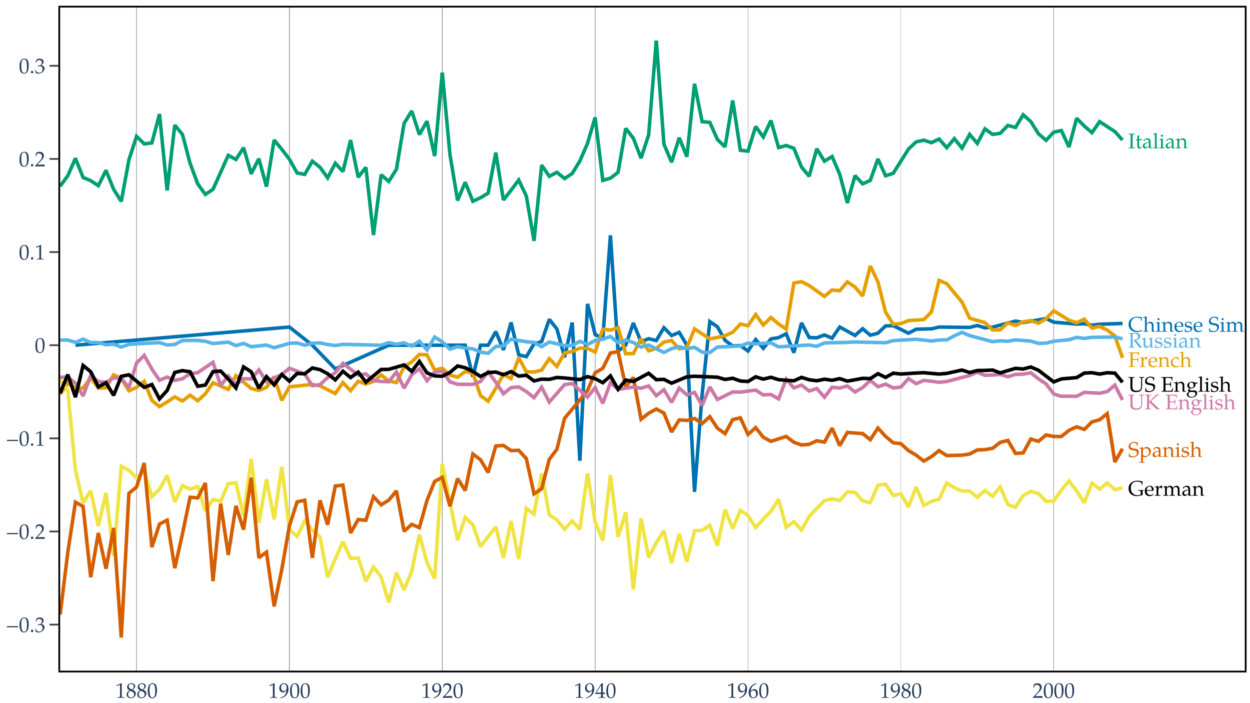
Note: General sentiment is based on the annual average projection of January-mentioning sentences' embeddings onto the positive minus negative January sentiment dimension. To define the positivity dimension, we average the difference in embedding for following tuples (and their translations to each language): [("january is good for society","january is bad for society"), ("january is mostly good","january is corrupt"), ("january positively impacts our world","january negatively impacts our world"), ("january helps the economy","january hurts the economy"), ("january benefits society","january damages society")]. Sentences are from the Google Books Ngram corpus and embedded using BERT. Bands represent 95 percent confidence intervals produced by subsampling.

A.4 Comparison with alternative text-based approaches

A.4.1 Dictionary-based approach

A considerably simpler and popular method than ours, counts positive versus negative words to measure sentiment (Zhou, 2018). One limitation of this approach is that it often misses the context and subtleties of language, which humans would quickly discern from reading words in sequence. In fact a major engineering feat of BERT is that its underlying neural network pays attention to longer sequences of words (Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, and Polosukhin, 2017). However, the dictionary-based approach may be a reasonable alternative due to its simplicity.

Figure 8: Sentiment toward finance using an alternative dictionary-based approach



Note: Dictionary-based finance sentiment is based on the annual average sentiment of finance-mentioning sentences. Sentiment for a sentence is net positive words in a sentence normalized by total positive and negative words in the sentence. Sentences are from the Google Books Ngram corpus and the positive and negative words are from Loughran and McDonald (2020) and Chen and Skiena (2014).

We use a list of positive and negative word for each language. For English we use the Loughran and McDonald (2011) dictionary. For all other language we rely on Chen and Skiena (2014). The sentiment for each sentence is the number of net positive words in a sentence normalized by total positive and negative words in the sentence. We aggregate the sentence sentiments, weighted by their frequency in a year, to get sentiment for the year. Based on the dictionary approach, we get a more volatile score, illustrated in Figure 8, and a non-significant relationships with disasters, reported in Table 9.

A.4.2 Alternative language embedding-based approaches

As mentioned, our language embedding approach builds on Kozlowski, Taddy, and Evans (2019), but differs in an important way. Kozlowski, Taddy, and Evans (2019) fit a word embedding model (e.g. word2vec, glove) to each decade of sentences. They then measure the

Table 9: Dictionary-based approach: Natural disaster effects on financial sentiment

	Finance sentiment growth$_{t+1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Natural Disaster$_t$	593.65	596.80	-139.39	595.86		
	(569.80)	(572.67)	(111.36)	(575.24)		
War$_t$		-52.24		-47.27		
		(104.78)		(98.86)		
Natural Disaster$_t$ $\times$ Low Insured$_t$			1094.32			
			(614.91)			
logKilled$_t$				-11.27		-14.58
				(14.72)		(19.50)
Drought$_t$					150.53	115.07
					(253.08)	(280.46)
Earthquake$_t$					1426.29	1435.65
					(1213.20)	(1236.52)
Epidemic$_t$					-71.98	-67.91
					(218.51)	(207.31)
Extremetemp$_t$					48.80	45.61
					(60.12)	(55.22)
Flood$_t$					-479.58	-483.40
					(260.25)	(265.39)
Landslide$_t$					-790.37	-816.91
					(1148.38)	(1188.96)
Storm$_t$					54.44	9.49
					(33.94)	(42.08)
Fog$_t$					-111.12	-130.96
					(145.37)	(135.32)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.17	0.17	0.18	0.17	0.20	0.20
Obs	851	851	851	851	851	851

cosine similarity once for each phrase of interest. The variation in their measures of culture come from variation in term frequencies but also from estimation error that generates variation in these fitted language models. By contrast, we use a pretrained language model (BERT), measure cosine similarity once for each phrase of interest, and then average these cosine similarities for each year (and language). Variation in culture in our approach is due only to term frequencies, as language model error is held fixed over time.

We attempt to apply the Kozlowski, Taddy, and Evans (2019) method by fitting three word embedding models, word2vec, glove, and fasttext, to every language-year in our panel. We find, however, that the finance sentiment series this approach generates are

Table 10: Word2Vec as an alternative model: Natural disaster effects on financial sentiment

	Finance sentiment growth$_{t+1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Natural Disaster$_t$	30.06	35.08	-1.07	35.02		
	(41.27)	(43.93)	(76.60)	(44.05)		
War$_t$		-93.86		-92.07		
		(61.62)		(61.79)		
Natural Disaster$_t$ $\times$ Low Insured$_t$			46.44			
			(96.54)			
logKilled$_t$				-7.41		-7.39
				(4.62)		(4.64)
Drought$_t$					120.48**	98.49
					(44.95)	(53.27)
Earthquake$_t$					-14.06	-9.57
					(52.08)	(52.22)
Epidemic$_t$					126.78	128.75
					(164.78)	(159.69)
Extremetemp$_t$					-82.12	-83.56
					(278.56)	(279.97)
Flood$_t$					153.86*	149.83*
					(75.71)	(74.62)
Landslide$_t$					173.61	160.23
					(135.12)	(137.62)
Storm$_t$					-52.45	-48.14
					(109.40)	(106.87)
Fog$_t$					28.81	25.49
					(92.94)	(89.12)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.17	0.17	0.17	0.17	0.17	0.17
Obs	833	833	833	833	833	833

highly noisy. In Tables 10, 11, and 12, we report the natural disaster regression estimates using these alternative text-based measures. The tables show that, as one may expect, the noisier measures generate considerable parameter uncertainty.

A.5 Severe disaster cutoff

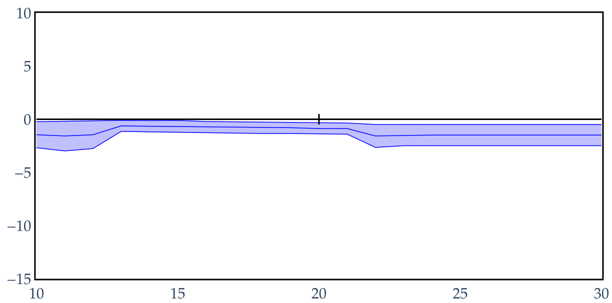
Figures 9 and 10 shows our estimates of the effect of natural disasters are mostly robust to varying the cutoff for the fraction of the population killed by the disaster. Lower cutoffs include more benign natural disasters, while higher cutoffs concentrate the treatment effect estimates on fewer but more fatal disasters. As a result, the point estimates for more fatal

Table 11: GloVe as an alternative model: Natural disaster effects on financial sentiment

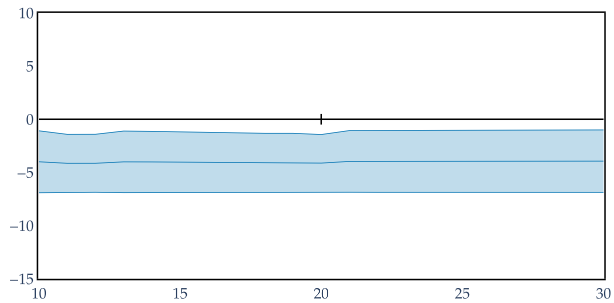
	Finance sentiment growth$_{t+1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Natural Disaster$_t$	600.28	596.84	1441.70	596.19		
	(472.00)	(473.03)	(1614.60)	(467.86)		
War$_t$		111.32		115.63		
		(151.28)		(151.56)		
Natural Disaster$_t$ × Low Insured$_t$			-1302.80			
			(1616.99)			
logKilled$_t$				-16.00		-14.97
				(17.79)		(18.47)
Drought$_t$					321.47*	260.52
					(147.45)	(158.39)
Earthquake$_t$					71.10	78.30
					(108.60)	(114.66)
Epidemic$_t$					83.13	61.36
					(118.79)	(141.98)
Extremetemp$_t$					4343.79	4341.38
					(4766.25)	(4764.87)
Flood$_t$					33.42	25.34
					(141.04)	(144.75)
Landslide$_t$					-124.93***	-151.91***
					(34.28)	(41.61)
Storm$_t$					-303.68	-292.19
					(257.38)	(298.14)
Fog$_t$					284.91	278.55
					(386.15)	(387.92)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.13	0.13	0.14	0.13	0.15	0.15
Obs	808	808	808	808	808	808

disasters are generally larger in magnitude and feature greater parameter uncertainty.

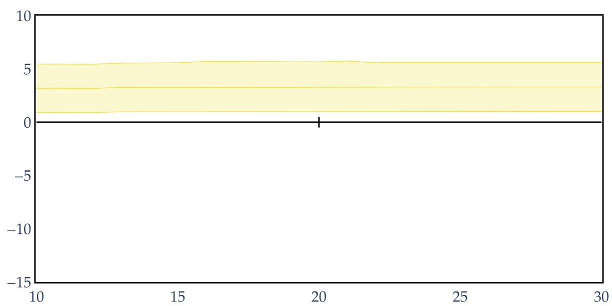
Figure 9: Robustness to the severe disaster cutoff for natural disaster groups



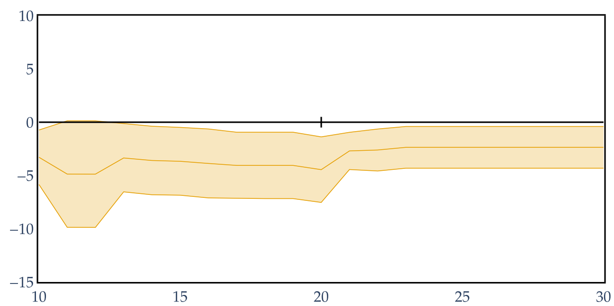
(a) Natural Disaster (any)



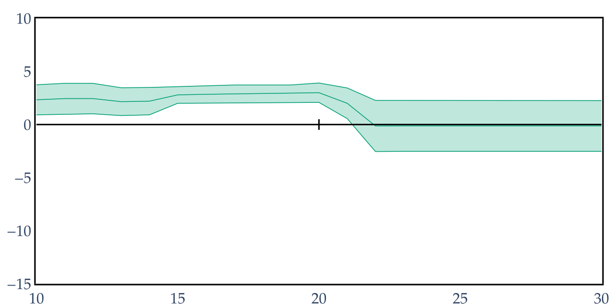
(b) Biological



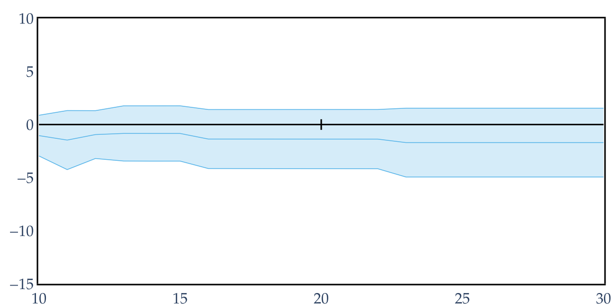
(c) Climatological



(d) Geophysical



(e) Hydrological



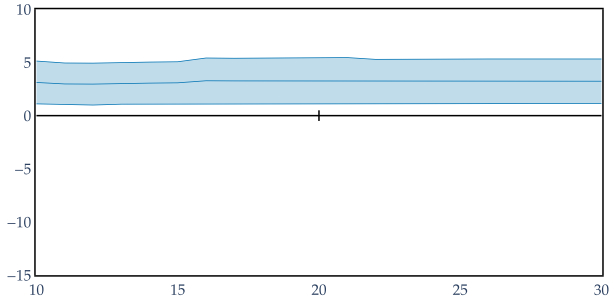
(f) Meteorological

Note: The figure shows how the estimated treatment effects of severe natural disaster groups change as we vary the minimum number of deaths per million for a disaster to be considered severe, thus filtering out less devastating disasters. Bands indicate 90% confidence intervals.

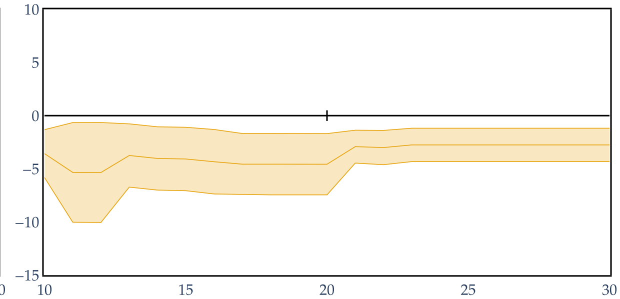
Table 12: FastText as an alternative model: Natural disaster effects on financial sentiment

	Finance sentiment growth$_{t+1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Natural Disaster$_t$	-34.41	-39.17	-103.05	-39.26		
	(81.37)	(78.75)	(72.71)	(79.01)		
War$_t$		88.20		90.97		
		(176.68)		(177.71)		
Natural Disaster$_t$ $\times$ Low Insured$_t$			102.32			
			(168.59)			
logKilled$_t$				-11.09		-10.89
				(12.51)		(12.69)
Drought$_t$					25.67	-6.74
					(199.08)	(216.77)
Earthquake$_t$					-13.88	-7.27
					(111.39)	(114.81)
Epidemic$_t$					176.59	179.55
					(139.59)	(147.85)
Extremetemp$_t$					-28.82	-30.92
					(115.34)	(116.23)
Flood$_t$					-260.87***	-266.80***
					(49.75)	(52.43)
Landslide$_t$					206.24*	186.51*
					(93.04)	(86.99)
Storm$_t$					204.43	210.79
					(187.13)	(211.38)
Fog$_t$					-6.41	-11.32
					(75.17)	(78.11)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.14	0.14	0.14	0.14	0.14	0.14
Obs	840	840	840	840	840	840

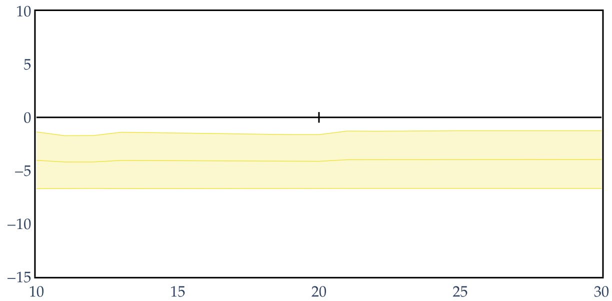
Figure 10: Robustness to the severe disaster cutoff for natural disaster types



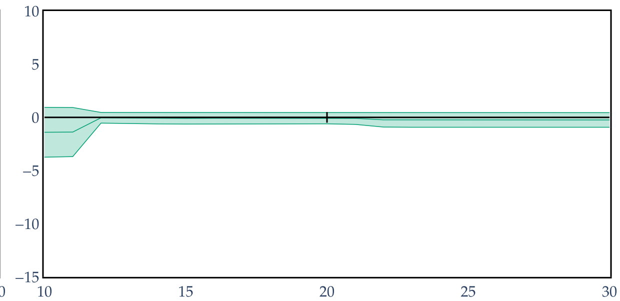
(a) Drought



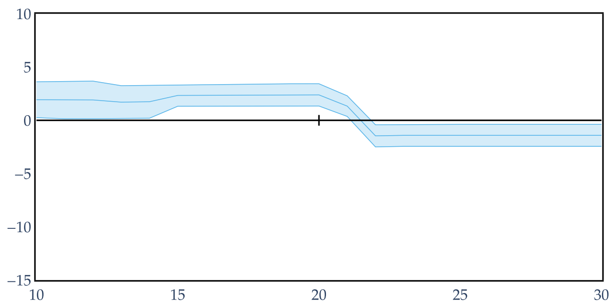
(b) Earthquake



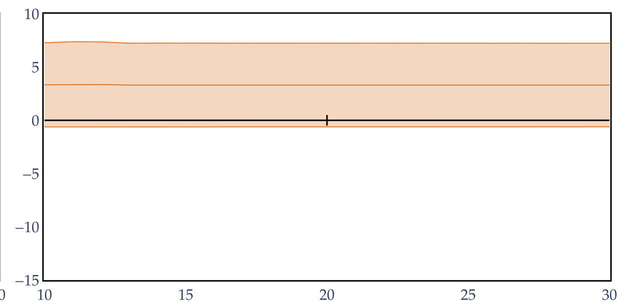
(c) Epidemic



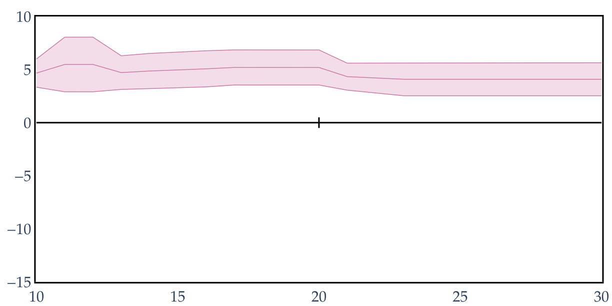
(d) Extreme Temperature



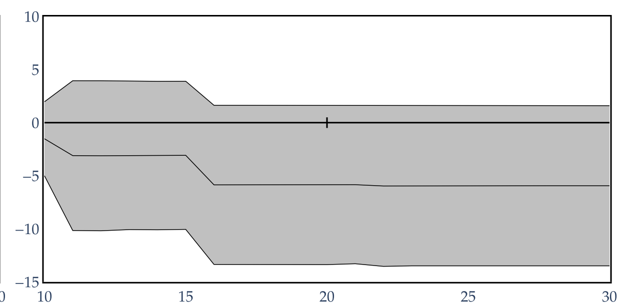
(e) Flood



(f) Fog



(g) Landslide



(h) Storm

Note: The figure shows how the estimated treatment effects of severe natural disaster types change as we vary the minimum number of deaths per million for a disaster to be considered severe, thus filtering out less devastating disasters. Bands indicate 90% confidence intervals.

Manish Jha

PhD Finance Candidate, WashU
https://mjha91.github.io

CB 1133, Simon Hall
St. Louis, MO 63130
+1 (314) 250-8129
mjha@wustl.edu

Education

Washington University in St. Louis
PhD in Finance, 2016–2021 (expected)

Indian Institute of Technology (IIT) Kanpur
Bachelors in Technology (Chemical Engineering), 2012

Research

Interests

Corporate Governance, Fintech, Machine Learning, Text Analysis

Job Market Paper

1. **Catching the Conscience of Kings: How Activists Pander Mutual Funds**
   Do hedge fund activists tailor their campaigns to pander to mutual fund families? And if so, does the strategy work? Using supervised machine learning on the fund family's proxy voting choices, I estimate their preferences. I find that activists align proxy communications with the preferences of fund families that own a larger share in targeted firms. Tailored campaigns enjoy improved shareholder attention, more votes, and a greater likelihood of success. Furthermore, activists learn from interactions with fund families and align their campaigns better in subsequent attacks. Targeted firms also align their management proposals to larger shareholders, specifically during the attacks. My findings suggest that activism helps push shareholders' implicit agendas.
   Presented at: Eastern Finance Association (scheduled), Georgia State University, Midwest Finance Association (scheduled), Stanford Rising Scholars Conference, University of North Texas, Virginia Tech, Washington University
   Media: Promarket

Working Papers

2. **Natural disaster effects on popular sentiment toward finance** with *Hongyi Liu*, and *Asaf Manela*
   We analyze the effects of natural disasters on popular sentiments towards finance. The sentiment is based on millions of books, spanning eight languages during the 1870–2009 period. Finance sentiment declines after epidemics and earthquakes, but rises following droughts, floods, and landslides. These heterogeneous effects of natural disasters suggest finance sentiment responds differently to the realization of insured versus uninsured risks. Using local projections, we find that positive shocks to finance sentiment have positive and persistent effects on economic growth. Our estimates predict a contraction in finance sentiment due to the COVID-19 pandemic that will exacerbate its long-term economic damage.
   Solicited by *Journal of Financial and Quantitative Analysis (JFQA)*

3. Does finance benefit society? A language embedding approach with *Hongyi Liu*, and *Asaf Manela*

We measure popular sentiment toward finance using a computational linguistics approach applied to millions of books published over hundreds of years. Our method generates sentiment scores for each year from 1800 to 2009 across eight languages: Chinese (Simplified), French, German, Italian, Russian, Spanish, UK English, and US English. We document persistent differences in finance sentiment across countries despite ample time-series variation. Using Amazon mTurk, we validate our machine learning results with native speakers from each country. The changes in sentiments match with historical events and could provide a glimpse into different financial setups across regions.

Presented at: Greater China Area Finance Conference, Norwegian School of Economics, Virtual Finance Seminar (Michigan State, UI Chicago), Volatility Institute at NYU Shanghai, Washington University

Media: Bloomberg, Promarket

4. Bonds Lie in the Portfolio of the Beholder: Do Bonds Affect Equity Monitoring? with *Todd Gormley*

We analyze whether the increasingly large bond holdings of institutional investors are associated with how actively institutions vote and monitor their equity investments. We find that institutions conduct more governance research and are less likely to follow proxy advisor vote recommendations for companies whose bonds represent a larger proportion of their overall portfolio. The findings are driven by bonds held in funds that primarily hold equity positions, particularly actively managed funds, and concentrated among proposals more likely to require an investor's attention. Overall, our findings suggest that institutions' bond holdings contribute to their incentive to be engaged monitors.

Presented at: Financial Research Association Conference (early ideas)

Work in progress

News Coverage and Mutual Fund's ESG decisions

Efficacy of Behind the Scene Engagements with *Todd Gormley*

Teaching

Teaching Assistant, PhD Level

Empirical Methods in Finance with *Radha Gopalan* (2019)

Evaluation: Mean 9.5, Median 10

Teaching Assistant, Undergraduate and Masters Level

Advanced Corporate Finance

Financing, *Radha Gopalan* (2017, 2018); *Mark Leary* (2018)

Mergers and Acquisitions, *Armando Gomes* (2020)

Valuation, *Radha Gopalan* (2019); *Todd Gormley* (2017–2021)

Asset Pricing

Investment Theory, *Thomas Maurer* (2017)

Options & Futures, *Thomas Maurer* (2017)

Volunteer Tutor

Each One Teach One, St. Louis Elementary School (2018)

HSBC Financial Skills Exchange Program, Bangalore (2016)

Non-Academic Employment

Hongkong and Shanghai Banking Corporation (HSBC)

Sovereign Bond Research, 2014–2016

Published trade ideas for sovereign debts of the Czech Republic, Poland, Hungary, Romania, Russia, Turkey, Israel, and South Africa. Developed frameworks to predict emerging market policy rates.

Reliance Industries

Options Trading, 2012–2014

Traded Nifty50 (large-cap Indian stocks) options on behalf of the treasury. Executed trade recommendations generated by algorithm trading systems.

Tata Steel

Mineral Processing Research, 2011

Conducted lab experiments to develop flow sheets for recovering iron values from high alumina slimes.

Miscellaneous

Citizenship: Indian (F1 Visa)
Languages: English, Hindi, Maithili

Programming

Courses: Deep Learning, Machine Learning, Natural Language Processing, Tensorflow, Text Analysis
Languages: C/C++, CSS, HTML, Java, JavaScript, Julia, Matlab, Mathematica, PHP, Python, SAS, SQL, Stata, Visual Basic
Cloud infrastructure: High Performance Computing, Software Containers

Certifications

Financial Risk Manager (FRM)
Chartered Financial Analyst (CFA), Level II
Institute of Actuaries of India: Mathematics & Statistics, Business Economics

Awards & Fellowships

GMAT waiver for PhD program application, 2016
Euromoney excellence award for HSBC research team, 2015
State rank one in all India engineering exam, 2008
Super30 scholarship, 2007–2008

References

Radha Gopalan
Professor of Finance
Washington University
+1 (314) 935-9196
gopalan@wustl.edu

Todd Gormley (Chair)
Associate Professor of Finance
Washington University
+1 (314) 935-7171
gormley@wustl.edu

Asaf Manela
Associate Professor of Finance
Washington University
+1 (314) 935-9178
amanela@wustl.edu

Last updated: February 2021